

Anti-apartheid moves to Mainz

This week's continuation of last year's protests at the presence of South African archaeologists at an international conference should remind the rest of us of unfinished business.

THIS time last year, there was a famous gathering at the University of Southampton that would have been the Thirteenth World Archaeological Congress if approval of the occasion had not previously been withdrawn by the International Union of Pre-historical and Protohistorical Sciences (IUPPS). That train of events came about because the Southampton organizers, responding as they said at the time to pressure from the local branch of the Association of University Teachers, the university's student union, the Southampton City Council and the British Anti-Apartheid Society, had decided to disinvite intended participants from the Republic of South Africa and Namibia. The official version of the congress is being held this week at Mainz in West Germany. German and other anti-apartheid groups have advertised in advance their intention to use the occasion to demonstrate their opposition to the free participation of South African scientists in international conferences of any kind. There are many in the scientific community who believe these protests to be justified. Many others are silent on the issue. The result is that great damage is done to the integrity of international science. Whether South Africa is much damaged, or the ending of apartheid hastened, are different matters.

Since the appearance of a report on South Africa and a discussion of the boycott issue earlier this year (*Nature* 327, 259 & 269; 1987), there has been a steady flow of correspondence on the question (see, for example, page 9, this issue), much of which has not yet been published. Readers will have noted, some perhaps with surprise, that both sides in this public argument share the common belief that the apartheid system institutionalized by the government of South Africa is intolerable and must be abandoned. The questions that have arisen and will arise, all of which depend on that common goal, are whether a boycott of South African scientists, or more generally of science, would speed the end of apartheid, whether the continued well-being of the intellectual community will help the post-apartheid society of South Africa (if there is one) and whether some kind of selective boycott would meet a wider interest.

Clamour

For the purposes of public clamour, the difficulty with South Africa is that it is much more complicated than is readily appreciated. If South Africa were likened to, say, an aircraft in which one group of passengers (all white) had by violence seized command and directed that twice as many fellow-passengers (all black) should do their bidding, the world would cry outrage and, while denying the aircraft fuel and landing-rights, would marshal its sharpshooters to kill the hijackers in such a way that their fellow-travellers would not suffer. South Africa is not like that. Apart from being a sovereign state, and also the source of much of the meagre prosperity of southern Africa, the theory that the hijackers are a gang that acts in concert is mistaken. What stands out from the correspondence so far published (there is more to come) is that the gang (none of whose leaders has written in) is riven. Maddening though it may be for the demonstrators at Mainz to acknowledge this, South Africa is crammed with heroes in the cause of reform, both people (many

of whom are in prison) and institutions. Three titularly white universities (Witswatersrand, Natal and Cape Town) have gone as far as prudence would dictate in defying the government (but seem ready to go further). Stellenbosch, traditionally an Afrikaaner institution, embodies a large group of academics seeking to break with the past. Is this the time to deny these places comfort, even help?

Boycotts

The other side of that coin is that boycotts can be effective. Decisions by other nations not to send to South Africa teams skilled at rugby football or cricket have directly stimulated changes of the apartheid rules (but most black South Africans have questions so much more urgent on their minds that they have taken only meagre advantage of the new arrangements). In science, the mixture of ostracism (disinvitation), discourtesy (not turning up) and personal abuse to which South Africans have been made accustomed has had a dramatic effect; people have been compelled to think out where they stand. (Many of those who say they have in the process discovered that apartheid must end have yet to appreciate the urgency of South Africa's position.) So why not go the whole hog, sever all intellectual links with South Africa and then wait for apartheid to collapse?

Because that would not happen. Instead, the republic's government would do what, by its own lights, it should have done years ago — emasculate or even shut down the dissident universities, imprison the academics who have nowhere else to go and await with an appropriate measure of unreality the cataclysm it has designed even for those who vote it into office (but many of them have organized bolt-holes in less inconvenient places). The demonstrators at Mainz, not to mention many other better-informed people, fail to recognize that these institutions are their hostages in the fight against apartheid. Rugby football never had that distinction.

This is the crux on which the issue of the boycott turns. The business community in South Africa, while just as anxious for reform, calculates the future on a shorter timescale. The valiant universities have three functions — today, to say what they believe; tomorrow, to change the way that other people think; for the future, to educate their generation's young. If these institutions survive the years ahead, they will surely be worth inheriting in the last of these roles. On a shorter timescale, it may be held that, in their eagerness to change the way that people think and behave, they have too often neglected the hard task of telling how to get from here to a better there; their neglectfulness will, of course, become habitual if their academics are uniformly ostracized. For now, what these institutions need is help, moral (to keep embattled spirits up) and financial (to teach more black students). Why are the demonstrators at Mainz and their supporters not contributing to those causes?

Does not that amount to a case for a selective boycott? That is how it sounds, but the argument is incomplete. Both the African National Congress in Lusaka and the several recently formed quasi-professional groups in South Africa which have recognized that intellectual life must be kept alive will, given a chance,



operate a system by which local validation of individuals' fitness to travel abroad will determine who turns up at what conferences. That is a nonsense because it diminishes the correlation between those free from boycott and the interest of what they may have to say. It also threatens the integrity of professional life by transferring responsibility elsewhere. And it is a recipe by which intellectual life would be dangerously politicized. Who would allow that participation in next year's meeting of the Federation of Experimental Biology Societies in the United States should be open only to supporters of the Strategic Defense Initiative (or sceptics thereof)?

Nobody can seriously wish to see intellectual life evolve in such a way. The trouble is, that is how it is likely to go if nothing else is done. But must the professional community outside South Africa necessarily be so passive, wringing its hands about the state of affairs it sees, but supposing that there is nothing it can do to help to change it? It is easy to complain that South African academics should do more to change the probably tragic course of events ahead of them, but if they and their institutions are indeed overwhelmed in the years ahead, will their fellows elsewhere be able to wash their hands of the affair and say there is nothing they could have done? On the other side of the coin that says that intellectual life transcends frontiers is the legend that no part of it can be forgotten. So why not instead make common cause with the recognizable outposts of rationality in the knowledge that, then, there would soon be more of them? □

Do there go I?

For those with a sense of history, last week's report on US engineering is a chilling read.

THE National Science Foundation probably had an inkling of what the outcome would be when it asked the National Academy of Sciences to say what changes should be brought about to make US engineering industry more internationally competitive. Even so, practising engineers will not be as pleased with last week's panel report (see page 5), with its firm declaration that US engineers must learn to do what their colleagues elsewhere have been attempting in the decades since the Second World War — to set themselves international standards of design and performance.

A little reflection will nevertheless show that neglect of this precept is part of what has gone wrong with the US economy in recent years. US exports of high-technology products languish because their quality is not as good as that of the Japanese and because they cost more than the Europeans collectively produce. A partial explanation is that the United States has for so long been a self-contained market that people find it hard to change the patterns of their thought. And while the old flair for fixing things survives, vaunting ambition is now unhappily constrained by too much disappointment, on the launch-pad and in the market-place.

A few historical reflections should chase away that thin excuse. In the decade from 1945, the British government busily sought to improve the competitiveness of its industries by arranging for comparisons with more efficient industries elsewhere, then mostly in the United States. The Anglo-American Productivity Council, one of the chief vehicles of this work, produced a much longer string of rude reports than the United States has yet seen. The central message was that British industry should learn to compete in design, efficiency and price with that of the United States. In the present comparison between the United States and (mostly) Japan, the gap in performance is less, although the faltering of US engineering education is comparable with that in Britain in the 1950s. After a lag of thirty years or so, people in Britain seem willing to learn the lesson, although the mechanisms have still to be put in place. Will as much time have to pass in the United States? □

Phoenix emergent?

The British Association talked bravely last week about saving British science. What might it do?

REPORTS (see page 7) that the British Association is about to do something for the advancement of science should not be too quickly discounted. Some of those at Belfast last week may have forgotten that the association came into being (in the year of the optimistic Great Reform Bill of 1832) when British science was also in a poor state. A century after Newton, the most venerable scientific institution of the time, the Royal Society, seemed to have run out of steam, and could certainly not compare in influence with the engineering and agricultural institutions or in the excitement of its proceedings with the Royal Institution (where Faraday was at work). The British Association remained an important influence in British public life for the remainder of the nineteenth century, until science was professionalized and disciplinary societies became the obvious means of communication. But latterly, the association has lost its way, not quite knowing what it has been for, and has been roundly criticized on that account — even, sadly, by *Nature*.

Could that now change? To be fair, the association by itself cannot solve its own dilemma. As when it began, people's expectations of what it might accomplish are more important than its officers' resolutions. At the outset, the crying need was for a means of informing an interested, educated and curious public about science and its applications. That remains the association's objective, even though other more efficient means (such as television) have emerged. Now, in Britain, the crying need is to defend the research enterprise against external pressures, financial (tightening budgets), ideological (the abolition of university tenure) and doctrinal (the involvement of industry). How could an open-membership organization help towards that end?

The urgent need is factional, of course. The research community is only a small part of British professional life and a smaller part of the British population. But, just as the British Association cannot be sufficient to itself, so the research community cannot hope for a successful defence of its interests without carrying its case to the wider community. Much of its weakness now stems from its loss, by a decade ago, of the popular understanding and enthusiasm that helped the British Association along in the first few years of its existence. Is there now a chance that the interests of those in research and those outside it can again coincide?

That is the sense in which last week's Belfast meeting could be an interesting starting-point for helpful innovation. During the past decade and more of near-stagnation, British researchers have fallen into the habit of complaining to each other, but quietly, as if fearful of being overheard. Now that it seems as if the government intends to act, and as if nobody's interest will be untouched, there are the strongest reasons why people should speak out with constructive criticism of what may be planned (see *Nature* 328, 745; 1987). That may account for some of last week's outspokenness. But, on its own, that will not suffice. If the British research community is to reassert its interests, it must build a constituency of support. Is it too much to hope that the British Association might be a means of doing that? Only if researchers are more ready than has been their habit to turn up to tell the world at large why what they are doing is interesting, even important. Then it would also be necessary that the association's stiff procedures should be changed so as to allow it to speak out as a representative organization, with all the risks that entails. The association could do worse than provide the low-pressure pressure group Save British Science with the organization it badly needs. It will be interesting to see whether the association has been able to grasp some of these nettles by the time of next year's meeting, which is at Oxford. That could be an even better meeting. □

Tobacco companies win round in health battle

- Smokers warned on packages
- Anti-smoking campaigns continue

Washington

IN a decision hailed by tobacco companies and damned by anti-smoking forces, a US appeals court in Boston last week ruled that cigarette manufacturers cannot be sued for damages from smoking on the grounds that they failed to warn consumers of the potential hazards. Since 1965, a warning label has appeared on cigarette packages identifying smoking as a potentially unsafe activity.

The decision arose out of a \$3-million suit brought by the heirs of Joseph Palmer against the Liggett & Myers Tobacco Co. of Durham, North Carolina. Palmer, a heavy smoker, died of lung cancer seven years ago. A lower court ruled in Palmer's favour, but the three-judge appeals court has unanimously reversed that decision.

This is not the first time courts have ruled that consumers are adequately warned about the dangers of smoking by the package labelling. Earlier last month, an appeals court in Atlanta made a similar decision in favour of the American Brands Co., as did a court in Philadelphia last year. The Philadelphia case was appealed to the Supreme Court, which declined to hear the appeal.

The Federal Cigarette Labeling and Advertising Act of 1965 required that cigarette packages contain the warning that smoking may be hazardous to health. In 1984, that message was changed to four more-specific warnings that change every three months during the year.

Opponents of smoking have argued that although tobacco companies comply with

SURGEON GENERAL'S WARNING: Quitting Smoking Now Greatly Reduces Serious Risks to Your Health.

the letter of the law by placing the messages on cigarette packages and advertising, their promotional campaigns convey the impression that smoking is glamorous rather than risky.

There are currently some 125 lawsuits pending against tobacco companies, seeking damages through product liability. A spokeswoman for Liggett & Myers says the Boston ruling will limit the scope of claims that can be brought. But according to Richard Daynard, a law professor at Northeastern University and chairman of the Tobacco Products Liability Project, there are many other avenues that litigants may pursue.

Plaintiffs can attempt to prove that cigarettes are an unreasonably dangerous

product, and that their risks exceed their benefits. Failure to list potentially hazardous ingredients — such as inorganic fibres like asbestos that may cause health problems — may also make tobacco companies liable for damages. Daynard also points out that many smokers now suffering health problems began smoking before 1965 when the warnings first appeared.

SURGEON GENERAL'S WARNING: Smoking Causes Lung Cancer, Heart Disease, Emphysema, And May Complicate Pregnancy.

The percentage of people smoking in the United States has been declining since the mid-1960s. Figures from the National Center for Health Statistics show that in 1965, 52.4 per cent of men over 20 and 34.1 per cent of women over 20 were smokers. In 1985, those figures had dropped to approximately 32 and 27 per cent respectively.

The National Cancer Institute (NCI) has identified a decrease in smoking as one of the key steps towards the goal of reducing cancer mortality by 50 per cent by the year 2000. According to Joseph Cullen, deputy director of cancer prevention and control at NCI, smoking is associated with some 30 per cent of all cancers.

As a biomedical research agency, NCI is prevented from lobbying directly for a smoking ban. But NCI is involved in clinical trials to identify effective approaches for reducing the frequency of smoking. Self-help, intervention by doctors or dentists, and media campaigns are all part of the constellation of programmes that Cullen believes have had an impact on 10 million people in 25 states. The next stage in NCI's plans is to establish regional coalitions to implement the tools NCI is developing to help people stop smoking.

Cullen worries that although smoking has been declining steadily, it will become harder to get people to kick the habit as reduction efforts get closer to a hard-core of chronic smokers. NCI has just launched a large-scale clinical trial on the health effects of heavy smoking (greater than 25 cigarettes per day) as part of an effort to find effective programmes to help heavy smokers to quit.

The other concern for anti-smoking forces is the growing popularity of smoking among poor and minority populations. Cullen says if things do not change, people in these populations will account for the largest numbers of cancer deaths.

Joseph Palca

Sulphur emissions cut

THE first binding international treaty on air pollution comes into effect this week. Under the Convention on Long-range Transboundary Air Pollution, it obliges 16 ratifying countries to cut back their sulphur emissions by at least 30 per cent on 1980 levels.

The United Nations Economic Commission for Europe (ECE) says ten countries have already achieved the 30 per cent reduction target, but these cutbacks have been partly offset by emission trends in other countries. Overall emissions of sulphur dioxide in Europe have fallen from 51.3 million tonnes in 1980 to 44.7 million in 1985. Data for North America show a decline from 27.8 to 24.5 million tonnes between 1980 and 1983. The ECE says further international efforts are needed to reduce emissions and decrease acid precipitation.

The implementation of the 30 per cent reduction is supervised by the Cooperative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe, through of 90 monitoring stations in 24 countries. K.J.

Supernova X-rays?

THE Australian space facility at Woomera, inactive since 1979, revived on 25 August with the launch of a West German X-ray telescope aboard a Skylark rocket. The brief flight, after which the telescope was retrieved and returned to West Germany, went ahead after protests from the local aborigine population had been overcome (see *Nature* 328, 104; 1987).

The instrument observed the recent supernova SN1987a in the Large Magellanic Cloud and preliminary data, indicating only weak X-ray emission from the supernova remnant, are now being analysed at the Max Planck Institute for Extraterrestrial Physics at Garching, near Munich. Joachim Trümper, director of the institute, hopes for continued collaboration, the next event perhaps being the launch of another X-ray telescope in 1988 to observe later stages of the expansion of the supernova remnant. S.D.

Shuttle test delay

PLANS to resume space shuttle launches in June 1988 were temporarily derailed on 27 August, when a series of technical hitches delayed and then postponed for two days the scheduled first test at Morton Thiokol's Wasatch Facility in Utah of the redesigned shuttle solid-fuel booster.

On the first firing attempt, a hose broke; on the second, there was a computer failure. D.L.

Launcher a success

JAPAN's space programme is progressing. Last Thursday the experimental three-stage H1 rocket successfully launched a satellite into geostationary orbit. □

Ariane set to go in two weeks, Columbus in the balance

Paris

EUROPE'S chances of a significant involvement in space in the next century could be radically affected by events in the next few months. As scientists at Kourou, the European Space Agency (ESA) launch site in French Guiana, start countdown rehearsal for the Ariane 3 rocket, following an 18-month delay, ESA officials are shuttling between Paris and Washington in an effort to complete negotiations with the US National Aeronautics and Space Administration (NASA) on Europe's Columbus contribution to the US space station.

A successful Ariane launch, set for 15 September, is imperative, both for ESA and for Arianespace, its French-based commercial arm. Recent demonstrations of Eastern interest in the commercial space market, with the Soviet Proton and Chinese Long March rocket launches (see *Nature* 328, 370; 1987 and 328, 659; 1987), have caused nail-biting among already tense European space scientists. Arianespace has a full order book, but it needs to carry out 8 or 9 launches a year until 1990 to make up for lost time. A turbo-pump failure in Ariane's third-stage rocket motor has interrupted Ariane-space's lead in commercial satellite launches since March 1986; a new motor had to be designed and tested.

The two telecommunications satellites, ECS4 and AUSSAT K3, due to be put into geostationary orbit by Ariane 3, have now been delivered to Kourou. The European satellite ECS4 will be used for international telephone, business and television communications, while AUSSAT K3 will provide a reliable voice and data link for Australian air-traffic controllers and will beam Australian television throughout the south west Pacific.

Meanwhile, the terms of an agreement allowing ESA to contribute a man-tended free-flying laboratory (MTFL) and a permanently attached pressurized laboratory module to NASA's planned space station have still not been settled. Earlier this year (see *Nature* 327, 180; 1987), the US Department of Defense upset negotiations when it insisted upon an option to use the space station for military research, a function specifically proscribed by ESA's constitution. Thanks to a standard diplomatic sleight of hand, whereby the crucial military use clause was relegated to an auxiliary agreement, a draft memorandum of understanding between NASA and ESA has been drawn up and, according to a Washington spokesman for ESA, "military use of the platform is not an issue in our discussions with NASA".

Approval of the memorandum must take place in time for the ministerial coun-

cil meeting of ESA on 8 and 9 November, when member states will be asked to commit themselves, both legally and financially, to the Columbus programme (comprising MTFL, Hermes and the pressurized module, as well as the British polar orbiting platform). "We are currently in the preparatory phase of the Columbus programme", said an ESA spokesman in Paris, "but we have to enter the development phase, and this will mean commitment and money for 10-20 years. The ministers have to decide yes or no, or yes with amendments."

It was said to be "quite likely" that member states will be asked to increase their contributions to the ESA budget at the November meeting. For this reason, Britain's stop-go attitude to its involvement with the space programme has

caused some concern at ESA. "We are following developments with great attention", said the ESA spokesman, "because Britain is an important member country and we would like the British to continue their efforts in space."

France is particularly keen to see a positive outcome at the November meetings but would favour maintaining the Columbus programme even without collaboration with NASA.

The coherence of the programme as a potentially autonomous unit will be the subject of an ESA Task Force report due in mid-September. The MTFL would be serviced at 6-monthly intervals by the (largely French) Hermes shuttle, and both would be launched from Ariane 5 rockets. Go-ahead for Hermes and Ariane 5 will also provide a boost to the economy of Guiana, a French territory, where a Hermes landing strip, a third rocket launch complex and Ariane 5 engine test facilities are planned.

Peter Coles

Storm in space-based teacup over Soviet rocket

London

PART of the British press has been occupied during the summer with the theme of Soviet 'industrial domination of space' and an inconclusive argument about the intended function of the new Soviet launching rocket Energiya. Soviet students of the British press may well think this excitement is a response to disappointments about government support for the British National Space Centre.

Three long articles in *The Times*, describing Soviet plans for space, referred in particular to the well-worn theme of how orbital power stations could collect solar radiation and relay it to the surface of the Earth so as to light Soviet cities and for other purposes.

This conclusion seems to have been taken sufficiently seriously by the Astronomer Royal, Sir Graham Smith, and Sir Bernard Lovell, to have provoked a warning of the hazards of such satellites. In a letter to *The Times* on 20 August, the radioastronomers say that the development of the Energiya launcher should be seen in military rather than civilian terms.

They also say that plans for collecting solar radiation are uneconomic by a factor of 100 and would be hazardous both because misalignment of the mirrors could harm people living near the ground receivers and by the atmospheric pollution arising from the need to launch one Energiya rocket every day. They add that solar-energy collectors driving infrared lasers could be directed at distant populations simply by steering the collecting mirrors, and call for international negotiation

in the context of the Treaty on the Peaceful Uses of Outer Space.

Lovell has long been an opponent of energy-collecting satellites, usually on the grounds that the microwave radiation suggested in earlier US schemes as a means of transferring solar energy to the Earth would interfere with radioastronomy.

A riposte from Mr Gerry Webb, managing director of the London-based consulting company Commercial Space Technologies, which claims to have provided much of the data on which the original articles in *The Times* were based, says that it is the "height of naivety" to interpret everything the Soviet Union does in space in military terms, that space power is "cheap, clean and limitless", that the "huge" capital costs that daunt Western planners are "more acceptable to the Soviet planning bureaux" and that this technology could "save the world" from "what every expert who has examined future trends" has concluded will be "inevitable industrial and economic collapse".

Neither side in this argument seems to have taken account of what is known of Soviet plans for the exploitation of space. The name of the new Soviet rocket, Energiya, is more probably a reference to its lifting power than to the functions of its payloads. Although there have been references, as for example by Academician Vladimir Kotelnikov, to the use of solar collectors a kilometre across, these probably indicate schemes for providing power to space stations and other structures in orbit.

Vera Rich

UK immigration authorities may use DNA fingerprinting

London

IMMIGRATION authorities in Britain are considering the introduction of the DNA fingerprinting test devised by Dr A. J. Jeffreys of the University of Leicester (see *Nature* 317, 818; 1985) as a routine procedure to establish paternity, depending on the results of a trial now in progress. But immigrants' associations are concerned at the social implications if the test becomes compulsory.

Britain's Home Office received 12,000 immigration applications last year from the wives and children of Bangladeshi and Pakistani men resident in the United Kingdom. The burden of proof to establish the family relationship is on the applicant, but this can be difficult because documentary evidence is often sketchy,

and at least one-quarter of the applicants are turned down. Blood tests can also be inconclusive, but DNA fingerprinting results are accepted as proof of paternity by the Home Office.

Immigration authorities are carrying out a trial of the genetic technique on 35 families from Bangladesh and Pakistan. A Home Office spokesman says the trial is designed to test the logistics as much as anything else, including the potential problems involved in obtaining blood samples abroad and sending them to Britain for the DNA fingerprinting. Results of the six-month trial are expected by the end of September.

The House of Commons select committee on immigration from the Indian subcontinent has recommended the

general introduction of DNA fingerprinting. The government's reply to the recommendation, published last week, states that the decision will be made once the results of the trial are assessed.

Although they welcome the potential of DNA fingerprinting to resolve individual disputed cases, immigrants' representatives are worried about the possibility that the technique will become routine procedure. In addition to questions about civil liberties, the representatives say there are also serious social implications for marriages in the highly conservative Asian

It's a serious matter, Sir, smuggling DNA fingerprinting kits...



Report urges US engineers to take an international view

Washington

US ENGINEERS must pay more attention to international activities to avoid being left behind in their rapidly changing profession. That conclusion comes from a report* of a panel of the National Academy of Engineering requested by the National Science Foundation (NSF). Although money will be needed for some of the panel's specific suggestions to develop this global perspective, the report's authors argue that a change of attitude is the key to ensuring the vitality of academic and industrial engineering.

The engineering world has changed in the past few decades. The United States has seen its balance of trade worsen, and now even high-technology imports exceed exports. Other countries have a proven capacity for technological innovation, and the world market-place for new technology has grown dramatically.

But the academy's report warns that reactionary attitudes exist within US engineering communities, notably a "bias against using what is 'not invented here'". That the National Technical Information Service spends only one per cent of its \$30 million annual budget collecting material from foreign countries is an example of this myopia.

At the same time, the panel says other countries have been quick to take advantage of engineering riches where they exist. As an example, the report cites statistics showing that in 1984 there were 13,160 Japanese studying at US universi-

ties, but only 730 Americans studying in Japan. The point is not to limit Japanese enrolment in US institutions, but to encourage more US involvement in education abroad.

To that end, the report urges the NSF to establish 100–200 postdoctoral fellowships annually for foreign study. The newly established NSF engineering centres should ensure that advances in other countries are not overlooked, and engineering schools should consider forming on-campus international technology assessment centres. The report urges engineering schools to require competence in a foreign language — with particular emphasis on Japanese and other Asian languages — as a requirement for graduation.

Albert R. C. Westwood, corporate director for research and development at Martin Marietta Corporation and chairman of the industrial panel of the academy's report, says engineers have traditionally been poor at communicating with foreign colleagues. But Robert White, president of the National Academy of Engineering, points out in a letter accompanying the report that companies and indeed nations "must be cooperative to be competitive".

Although concern has been expressed about technology transfer from the perspective of national security, the report argues that, with thoughtful implementation, "the benefits of international cooperation in engineering and technology are likely to outweigh risks". Protectionism is ultimately futile, "since technology inevitably diffuses".

Joseph Palca

societies if the fingerprinting proves non-paternity.

British immigration law requires the family relationship to be proved "on balance of probability", not beyond reasonable doubt, and immigrants' spokesmen say that documentary evidence can meet that requirement. They argue that DNA fingerprinting should be used as a last resort, rather than as a routine procedure, and they say the Bangladesh government has already objected to the principle of compulsory tests.

The genetic technique has already been used successfully by hundreds of immigrants in the past two years, through Dr Jeffreys.

Recently, Cellmark Diagnostics, a subsidiary of Imperial Chemical Industries (ICI), has taken over the testing work, under exclusive licence. Royalties from the fee of £105 per person go to the Lister Institute for Medical Research.

Cellmark's director, Philip Webb, says the laboratory could cope with increased demand from the Home Office at a negotiated price. The Home Office says that if the fingerprinting did become "a department tool", the costs would be met by the government.

Jeffreys did not foresee the potential use of DNA fingerprinting for immigrants, but with Britain actively considering its introduction, and Canada also interested, this may become one of the technique's most common uses.

Kathy Johnston

* *Strengthening US engineering through international co-operation: some recommendations for action.* National Academy of Engineering, Washington DC 1987.

Hepatitis B vaccine contract goes to untested product

London

SMITHKLINE Biologicals (SKB) and Merck Sharp and Dohme (MSD), two producers of the world's first genetically engineered yeast-derived hepatitis B vaccine which is billed as cheaper and more easily available than the old plasma-derived vaccine, have lost a major international supply contract for the vaccine to a small but growing South Korean pharmaceutical company.

The Korea Green Cross (no relation to its Japanese namesake) was last month awarded a contract to supply 425,500 doses of its locally manufactured plasma-derived Hepavax B vaccine, over five years, to Indonesian infants, at \$0.95 a

dose, in the first of a series of pilot vaccination projects planned for South-East Asia, Africa and Latin America.

The tender was a brainchild of the US-based International Task Force on hepatitis B, a loose affiliation of nine of the world's leading hepatitis experts that operates through an independent Seattle-based charity. At a recent meeting in London, Professor Palmer Beasley, a task-force member, pointed out that "with the AIDS pandemic, WHO [World Health Organization] is not in a position to address the rising incidence in the numbers of hepatitis B carriers and cases".

Frustrated by the lack of progress in

curbing the spread of the virus despite the availability of a vaccine since 1981, these experts, led now by Dr James Maynard, late of the US Centers for Disease Control, have sought to force down the cost of the vaccine to around \$1 a dose, to help developing countries to provide mass public-sector immunization programmes.

Price has certainly been a major stumbling block in many parts of Asia and Africa where one in ten are carriers of the hepatitis B virus. Of these, about half (2 million) will die from liver cirrhosis and liver cancer. China, Taiwan and South Korea all produce plasma-derived vaccines locally to meet domestic demand. None of these vaccines has been tested for safety and efficacy by an independent regulatory authority.

Health professionals now worry that in their zeal to get the vaccine to countries that need it, at a price they can afford, the task force may be endorsing a potentially substandard product. In evaluating the different bids, Richard Mahoney, executive secretary of the task force, said: "We did not wish to act as a regulatory agency and we did not question manufacturers' data on vaccine safety and efficacy". The task force also took the unorthodox step of declaring all vaccines acceptable if approved by the national regulatory authority of the country of origin. A spokesman for WHO confirmed that the Korean vaccine Hepavax B does not yet have the WHO seal of approval.

Although the Korea Green Cross has sold more than ten million doses of its vaccine in South Korea in the past five years, other vaccine manufacturers are worried that the company had produced only two batches of the special paediatric dose designed to meet the tender specification. Western regulatory authorities demand the manufacture of a minimum of five consecutive batches to establish vaccine stability before product approval can be granted.

Such is the demand for a cheap hepatitis vaccine in many parts of South-East Asia that some countries are willing to take a gamble rather than wait for the major companies to lower their price. Although prices of the new genetically engineered vaccine, at \$15 per dose, are half the price of the older plasma-derived type, they are still too high to make mass vaccination programmes feasible. It is hoped, however, that high-volume orders will lower the cost of the recombinant vaccine still further.

Belgium-based SmithKline Biologicals, whose rDNA hepatitis vaccine, Engerix B, was licensed in the United Kingdom this week, "pledges a commitment to the philosophy of low-cost hepatitis vaccines, provided that the volume of orders is sufficiently high", states Engerix B project manager Robert Hackett.

Miriam Ryan

Why the world needs population biology to solve its problems

Washington

CONSERVATIONISTS are sharpening their arguments. During the past month, pressure from Washington-based groups has led the World Bank to withdraw support from African cattle-ranching projects that would cause long-term economic harm and the Agency for International Development to publish lists of its own projects that stress short-term economic benefit at the expense of the environment. Now it is the turn of the Club of Earth, a pressure group of senior academic ecologists, to issue a statement stressing that support for population biology makes economic sense.

The club, whose members are drawn from the National Academy of Sciences and include G. Evelyn Hutchinson and E.O. Wilson, point out that it is population biology that solves "practical questions about maximum sustainable yields for fisheries, control of agricultural pests, containment of epidemics..." and that basic research is needed to keep the environment in shape just as much "as basic research in molecular biology is needed for treating human ills".

There the comparison ends. Financial support for biomedical research in the United States is more than two orders of magnitude higher than support for population biology. The major funding agency for population biology (taken to include ecology, ecosystem studies and systematic biology) is the National Science Foundation, which spends approximately \$60 million a year. It has "recently been increasing funding faster than for molecular biology" but still not fast enough to satisfy many ecologists. Jared Diamond, a Club of Earth member from the University of California at Los Angeles, points out

that most ecology graduate students have to leave the field and that any job opening will attract around 300 applicants.

Better knowledge of population biology could have helped avoid the collapse of the Peruvian fishery and the Californian sardine fishery, according to the Club of Earth. The former alone lost the world fishmeal worth \$1,000 million a year.

The evolution of malarial resistance, the seriousness of the AIDS (acquired immune deficiency syndrome) epidemic and the likely impact of genetically engineered organisms on the environment are all areas where failure to support population biology is preventing serious answers from emerging. Lack of taxonomists prevents the cataloguing of potentially valuable species, particularly in the tropics.

Diamond points out that as much as one-third of tropical biomass circulates through ants and termites but there are perhaps only six people in the world capable of classifying them. Cornell University's Tom Eisner, another club member, believes that failure to research the chemical interactions of plants, animals and fungi is slowing the development of more intelligent methods of pest control and will, in the long run, harm agriculture.

The kind of new perspective the club wants will not be easy to achieve and is unlikely to appear just by tinkering with research grants. Rather, there has to be an acceleration in the changes that are now occurring in political organizations such as the World Bank alongside a wider recognition of what population biology could contribute. To achieve that will require skilful management of Washington's complex political ecosystem — a fair intellectual challenge for the club's academics.

Alun Anderson

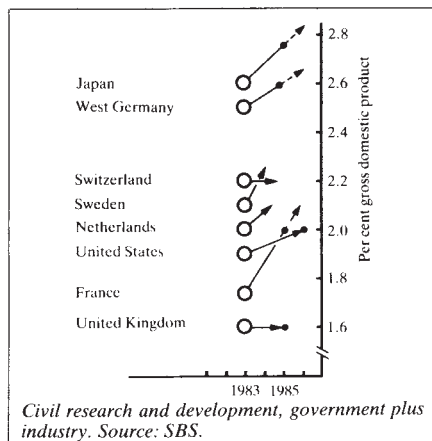
BA asserts itself in the defence of British science

Belfast

THE British Association for the Advancement of Science (BA) seems to have rediscovered its role as the defender of British science. At their annual meeting last week, the BA's officials announced that they would be keeping a close watch on the performance of the government's new committee, the Advisory Council on Science and Technology (ACOST), which they consider crucial for the future.

To earn the support of the scientific community, the BA says, the new arrangements will have to lead to real improvement in the factors that have led to such a decline in the morale of the scientific community in recent years. This means the government must "provide additional resources for scientific research and development", and also demonstrate real ministerial support for science and technology at the highest level.

Recent developments, including the resignation of Mr Roy Gibson as director of the British National Space Centre, are not encouraging, says the BA's president, Sir Kenneth Durham. That incident high-



lights a difference of opinion between the government and its critics about the respective roles of government and industry in supporting the two arms of research and development. One reason for the government's reluctance to set aside extra money for space research is said to be disappointment that industry is not itself spending more on research. But in the view of companies such as British Aerospace, some of the important work is too risky for the companies to undertake — and too dependent on decisions taken by international bodies such as the European Space Administration. Such research is exactly that which government should be supporting, in the BA's view.

Defence spending will also be monitored closely. The government has not committed itself to cutbacks, only to "review defence R&D spending". Durham holds out little hope that such a review will lead to much. For one thing, research is a relatively small component of the defence budget, and after inter-service wrangling there may be little left in the research column to be redistributed to the civil side.

The BA's determination to monitor the government's actions is an extension of the 'Science Audit', a procedure now in its third year, which started as a cooperative venture with government to gather an accurate picture of exactly how much was being spent on research, and to what effect. This year in Belfast the Science Audit banner has been used to give a platform to the views of Save British Science (SBS), which points out that Britain's expenditure on civil research and development, as a percentage of gross domestic product, is far lower than that of its main rivals (see graphic). Ironically, the BA was originally founded, in 1831, largely out of discontent at the state of British science. A similar discontent led to the founding of SBS in 1986 and seems also to be sending the BA back to its roots.

Charles Wenz

More condemnation of UK science policy

Belfast

THE pressure group Save British Science (SBS) has made its mark here. Dr John Mulvey, secretary of SBS, says the British government's new policies are but "manoeuvres designed to cope with inadequate funding".

In particular SBS wishes to see:

- National investment in civil research and development increased to 2.5 per cent of gross domestic product within 5 years.
- The University Grants Committee's side of the dual-support system restored and research council budgets increased to enable them to 'target' exploitable topics.
- Industry encouraged to spend more on civil research by effective measures — tax credits for instance — not simply by exhortation.
- The broader integrated science curriculum in schools — a welcome development — supported by the money to provide enough teachers and equipment for it to succeed.
- Government plans for a greater than 4 per cent increase in the proportion of young people graduating from universities and polytechnics by the year 2000.

The SBS view, in short, is that Britain's record in research has been, and still is, outstanding, but that the onus is on the government to turn this success to industrial and economic benefit. Charles Wenz

AIDS tests compulsory

London

THE Supreme Soviet has taken powers to authorize obligatory tests for AIDS (acquired immune deficiency syndrome), according to a statement by TASS, the Soviet news agency, a week ago. Tests will be administered to both foreigners and Soviet citizens when there are grounds for believing that people have been infected with the virus. Foreigners and stateless persons refusing to be tested when asked are liable to deportation under the regulations. Anybody deliberately exposing others to the risk of infection may be imprisoned for up to 5–8 years.

According to the president of the Soviet Academy of Medical Science, Dr Valentin Pokrovskii, on 17 August, there are some 130 people in the Soviet Union infected with AIDS, most of them foreigners. The only Soviet citizen with AIDS (out of 19 carriers) is said to be responding to treatment with two Soviet medications. In Bulgaria, testing for AIDS has been made compulsory for all newly-weds and pregnant women.

V.R.

Kidney problems

Bangalore

TRADING in human kidneys has got out of hand in the Indian metropolis of Bombay, prompting the state of Maharashtra Medical Council (MMC) to set up a special panel to root out illegal practices in kidney transplants.

Because cadaverous kidney transplants are not performed in India, trading in live human kidneys has become big business in Bombay. Entrepreneurs will offer a package deal on transplant services — a kidney and the necessary surgery — for prices ranging between Rs100,000 (£4,800) and Rs1.5 million (£72,000).

Advertisements will appear in newspapers both soliciting and offering live human kidneys. Typically agents hunt for poor migrant labourers willing to give up a kidney for a mere Rs3,000–8,000. Kidneys are normally sold through touts and doctors. Rich patients, often from Arab countries, will travel to India rather than pay more for the same procedure in the United States. Late last year, the Maharashtra state health minister reported that large numbers of foreigners are undergoing transplants in some top private hospitals. In Bombay alone there are about 10 kidney transplants each week.

The new panel will seek strict implementation of the Maharashtra Kidney Transplant Act of 1983. Once formally constituted, the panel will interview prospective donors and recipients, to protect the legal rights of both. The panel will include a nephrologist, a urologist and a lawyer. The state medical council favours setting up an institute equipped to store the harvested kidneys.

R.R.

Winds of change stirring for West German telecommunications

Munich

POWERFUL forces are moving into position in anticipation of a battle this autumn to deregulate portions of West Germany's telecommunications industry. Although the Deutsche Bundespost — the government agency in charge of post and telephone services — has so far successfully resisted efforts to break its monopoly over these services, an official commission has recommended that a "new age" for telecommunications be ushered in by the end of next year.

There has long been pressure on the Deutsche Bundespost to open up its markets to West German and international competition. But the Ministry for Posts and Telecommunications, which runs the Bundespost, has always had the power to resist. This is in part because of its tremendous size — with an annual investment of DM 17,000 million, it is West Germany's largest investor — and in part because it provides almost DM5,000 million in revenue every year.

But a commission, made up of representatives from industry, trades unions, government and political parties, calls for a separation of telephone and data-transmission from postal services and a new tariff structure allowing long-distance telephone rates to fall and local rates to rise on the basis of the cost of providing these services. Local telephone services, as well as long-distance networks, would remain under the administration of the 'postal monopoly' whereas most other services, such as telefax or teletext, would be opened up to West German and foreign suppliers over the next five years. Although hardly revolutionary in their scope, the commission's recommendations were met with withering criticism from the Social Democratic Party (SPD), the postal union and segments of the electronics industry.

The coupling of local with long-distance telephone services, and of postal delivery with telecommunications, has allowed the Bundespost to generate huge profits while still providing social services such as cheap package delivery and low local telephone rates. According to a survey by the Austrian magazine *Profil*, West Germany has only the sixth-highest local rates among Western European countries, but by far the highest costs for long-distance calls over 100 km, with rates three times as high as Britain or Belgium.

Supporters of the postal monopoly argue that the controlled prices are necessary for the Bundespost to provide the social services, such as cheap newspaper delivery, that are now expected of it. Furthermore, they say, the average

citizen profits from low local rates whereas big companies foot the bill for long-distance communication. A bizarre alliance of the left-of-centre SPD, some trades unions and the far-right Christian Social Union has publicly denounced the proposed deregulation because it opens up the market to companies that would "pick the raisins out" of the telecommunications industry — namely lucrative high-speed data and voice transmission between large cities — leaving the average citizen to pay the cost through higher rates or poorer service.

On the other side of the debate, the Economics Ministry and the Free Democratic Party argue that the current system is rife with inefficiency. They go so far as to favour the creation of competing satellite or cable networks as a way of forcing down long-distance prices.

Three of the most influential science and education organizations in West

Germany, the Wissenschaftsrat, West-deutsche Rektorenkonferenz and Deutsche Forschungsgemeinschaft (DFG), have added their voices to the current cacophony by calling for reduced rates for scientific use of telephone and data lines.

Wissenschaftsrat (science council) official Kurt-Jürgen Maass would prefer to see the telephone system administered like a private business, as the commission proposed, in the hope that scientists and students would receive the same discounts as they do when they buy computer hardware — up to 40 per cent in some cases.

The battle over deregulation will hot up after 16 September, when the commission report is officially submitted to Chancellor Helmut Kohl. A parliamentary resolution is expected to follow. The commission's target date for implementing changes in the Bundespost — 1 January 1989 — seems decidedly over-optimistic in the light of the political infighting that is bound to break out with the return of the Bundestag to Bonn after the summer recess is over.

Steven Dickman

Czech uranium miners up in arms over working conditions

London

URANIUM miners at the Zadni Chodov complex in West Bohemia (Czechoslovakia) have signed a "comprehensive complaint" against working conditions and the suppression of information about health hazards, according to the Czechoslovak party newspaper *Rude Pravo*. The "complaint" was signed several months ago; three reporters from *Rude Pravo* met with the principal petitioners in May. The delay in publication suggests high-level consultations before the article could appear.

The miners said that they had never been officially informed about the regulations on radiation hygiene, and only found out about them "by chance". One who complained about the high level of radioactivity to which miners were exposed was informed by the chief engineer that he was "talking too much". Miners who had to be transferred from underground work on health grounds suffered losses in pay and were not informed of the social security benefits to which they were entitled. Moreover, although their impaired health was due to past breaches of the safety regulations, and what they considered to be the management's "irresponsible gambling" with their health, they had difficulty in obtaining the documents necessary to file a claim for damages.

Doctors, the miners claimed, were "under the influence" of the management. Records of dosimetric measurements were

forged, and miners who asked for sick-leave after minor accidents were pressurized not to do so, because this would allegedly deprive the managers of their bonuses. One one occasion, the miners claimed, a ventilation system had been taken out of operation to save energy, although the mine was fully manned.

Representatives of the management, foremen and "other miners" interviewed by *Rude Pravo* denied these allegations, although the director of the enterprise did admit to a certain "disorder" in record-keeping. Management, he promised, would "deepen the methodological supervision of the entire mine" and had already instituted "consultation days", when workers could talk to specialists on matters of safety, wages and legal rights.

The protesting miners, however, say that they have lost all trust in the management and are "no longer interested" in talking to it. The *Rude Pravo* team noted that there was "no proof" of falsification and misreporting alleged by the protesters — but stressed that the incident revealed an "under-estimation of daily genuine contact with the people" on the part of the party and trades union organizations. Unresolved conflicts and lack of "openness", they concluded, create scope for misunderstandings, slanders, suspicion and accusations, to say nothing of a considerable expenditure of time, money and human effort in resolving such conflicts.

Vera Rich

Scientific boycott of South Africa

SIR—Whatever the merits of the original article on the boycott of South Africa that appeared in *Nature*, the path advocated in the polemic of J.G. Wilson (*Nature* 328, 288; 1987) would ensure that the Nationalist government remained in power for at least another four decades.

South Africans have been subjected to boycotts for long enough now for a reasonable evaluation to be made of the effects. The sports boycott is generally assumed by the international community to have been successful in changing government policy to interracial sport. It has also now convinced voting South Africans that any such reforms will not produce a positive response from the international community, and has removed the incentive for reform in other areas. Similarly, the oil and arms boycotts have resulted only in South Africa now being partly self-sufficient in these commodities and more immune to international action. The outcome of the South African elections is convincing evidence that the current flurry of disinvestment has not forced a change for the better in attitudes.

Recent history has shown clearly enough that the Nationalist government wields its power through education and economics. Change can be wrought only through altering the balance of this power in favour of the black population. The change in education could be achieved only by a massive investment in black education, particularly in teacher training. At the moment, there are not enough teachers and facilities to provide even a primary education to all black children. This has serious consequences for institutions such as the University of Cape Town where efforts to increase the number of black students admitted, despite government opposition, may be thwarted because of the inadequacies of the school system.

Education cannot be separated from economics; so much is clear from the British experience. As with other areas of South African life, black education suffers most when the economy is in decline. In addition, black labour was not well organized previously because black trades unions were illegal. Now that they are legal, the international community is ensuring their ineffectiveness by attempting to destroy their economic base through boycotts. Rising unemployment and rising illiteracy do nothing for the growth of power amongst the disenfranchised. Arguments that their lot could not get worse ignore reality. If you cannot afford bread, you certainly cannot afford books.

The solution is simple, if radical: a

massive investment in the South African economy. No doubt this will help the rich, but that is a small price to pay for the crucial benefits it would have for the growth in the economy and education of the underprivileged and disenfranchised people of South Africa. An investment in South Africa is an investment in the possibility of a nonracial future. Disinvestment simply gives the Nationalist government a licence to continue in power, a message reinforced by the results the South African election.

Voting South Africans have lived under threats and sanctions for a great many years now and have learned to live with them. They also see a certain inequality between the international treatment of South Africa and that of other nondemocratic states, which translates into further hardening of attitudes. An informal academic boycott has been in progress for some time and the efforts to formalize it cannot be expected to coerce a president who never went to a university, and will have little impact on the voting public. Communication is a quintessential facet of science and of academic research in general. Threats have not changed attitudes for the better, so why not at least apply the principle of communication to South Africans of whatever colour or persuasion?

KEVAN A.C. MARTIN

University College,
Oxford OX1 4BH, UK

SIR—The fact that J.G. Wilson (*Nature* 328, 288; 1987) did not like what he read in John Maddox's report of his visit to South Africa does not make Maddox *ipso facto* a racist, nor is the anti-apartheid cause furthered by a passionate misrepresentation of facts.

Far from young South African blacks being denied an education or careers in science, every effort is being made to encourage them to enter the science and engineering fields. Reflecting changes taking place, over the past two to three years the majority of winners of the South African Science Olympiad, a competition for pupils in their last two years of school, have been blacks. The Council for Scientific and Industrial Research tries to stimulate an interest in science careers by arranging visits of black and white school-leavers to its research laboratories, offers bursaries for both undergraduate and postgraduate study, and posts in research institutes on graduation on an equal-opportunity basis. Some large companies specifically sponsor black students on science and engineering courses at university, with guaranteed employment, again on an equal-opportunity basis, on completion of studies.

Rather than promoting a policy that would remove such backing and destroy jobs, or cause a mass exodus from South

Africa (where to and financed by whom?), surely a constructive action would be to build on the change which is taking place whether Wilson likes it or not. If he does have any real concern for black South Africans, Wilson could for one thing direct his energies towards campaigning for funds desperately required to build the one school a day until the year 2000 needed to keep pace with rapid population growth, thus ensuring the important role of blacks in science in the South Africa of the future.

PEGGY BEEZHOLD

35 Kensington Court,
London W8, UK

SIR—There was one major omission from your review of South African science (*Nature* 327, 269; 1987), which was the effect of the financial crisis, particularly the steep devaluation of a couple of years ago, on the funding situation. Political and social problems are, in the long run, probably the decisive ones, but in the short run, views of them are very subjective. There is nothing subjective about the financial situation.

For people like myself, who had intended to return to the United Kingdom on retirement, the exchange rate is nothing short of disastrous. For those who have no contacts outside South Africa, and therefore depend only on internal finance, the effect can be quite small. For scientists in general, who must obtain books and scientific supplies from overseas, or who depend on anything imported, the effect is very serious.

You discuss the effect of possible sanctions on the supply of journals. It seems to us at present much more likely that scientific journals will be priced right out of our reach than that they will be politically impossible to obtain.

On the personal level, selling a fairly large house in Durban is likely to provide enough funds to buy only a very moderately sized car in the United Kingdom. Indeed, my house cost me about £15,000 when I came here 15 years ago, and I expect to get about the same for it now.

It would be interesting to know who was the American who got R 75,000 working for the Council for Scientific and Industrial Research. I thought that in the public service only politicians got anything like that pay. University staff certainly do not.

DAVID A. H. TAYLOR

Department of Chemistry
and Applied Chemistry,
University of Natal,
King George V Avenue,
Durban 4001,
South Africa

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only. □

USA and poverty

SIR—In his review of Asimov's book *Past, Present, and Future* (*Nature* 327, 667; 1987), Michael Spencer suggests that Asimov's view of the world population problem is naive.

In the course of his own brief observations on the subject he raises the related question of poverty in developing countries, asking: "Why are so many Third World countries unable to rise out of poverty? Could it be relevant that the United States has only a twentieth of the world's population but accounts for one-third of its consumption of resources?"

The implication is that the United States is somehow responsible for the poverty of developing countries. Let us explore this direction further. Let us suppose that an enormous politico-geological cataclysm were to submerge the United States beneath the sea and leave the rest of the world unscathed. How would this event affect the third world?

The tone of the review leaves little doubt that, freed from the shackles of US exploitation, the developing countries would rise out of poverty.

More probably the opposite would happen. In the absence of the greedy, spendthrift Americans, who had hitherto accounted "for one-third of [the world's] consumption of resources", demand for these resources would drop by one-third. World-market prices of raw materials and crops would collapse, a major source of sustenance for third world countries would be withdrawn and, far from rising out of poverty, they would sink into deeper poverty.

JOCELYN TOMKIN

Department of Astronomy,
University of Texas,
Austin, Texas 78712, USA

What Fisher meant

SIR—It is possible to demonstrate R.A. Fisher's support for almost any idea by changing what he wrote, and Sibly's attempt to adduce his posthumous support for "the optimization methods underpinning the modern evolutionary approach to biology" is breathtaking in its simplicity: just change *genotype* to *gene* in Chapter II of *The Genetical Theory of Natural Selection*² because that is what "Fisher must have meant".

Taking such a liberty reveals a failure to understand *The Fundamental Theorem of Natural Selection* (the title of the chapter in question), whose point is precisely that in a diploid organism it is genotypes rather than genes that have to be considered. Fisher shows that the rate of increase in the mean fitness of a population is then equal to that component of the total variance which we nowadays call the additive genetic variance or the genic

variance (a partitioning of the total variance which would be redundant if genes, and not genotypes, were being considered).

The theorem is, admittedly, not an easy one, and is frequently misunderstood. Thus Turner³ has recently misrepresented it as relying on "the reductionist assumption that the effects of all genes on the variance could simply be added, which is, in effect, the assumption that there is no interaction between genes, not even the interaction between alleles which we call 'dominance'". Fisher made no such assumption (which, if made, would again be tantamount to dealing with genes and not genotypes); it is a *consequence* of the theorem that the dominance variance is not involved, not an assumption.

It sometimes seems as though "the modern evolutionary approach to biology" is so preoccupied with supposed maximization principles that in order to apply them it is prepared to omit some of the elementary facts of genetics, such as mating or dominance, as though Fisher had never written *The Genetical Theory*. The book quoted by Sibly⁴ is a good example, even Sewall Wright's idea of a "selective landscape" being misappropriated to describe a graph in which the independent variates are resources allocated to growth and to reproduction rather than the gene frequencies specified by Wright.

Biologists who import the theory of mathematical economics into evolutionary studies should at least avoid using the terminology invented by those who had supposed that the theory of mathematical genetics was more relevant, and they should resist absolutely the temptation to change words in quotations from the latter theory in order to match the preconceptions of the former.

A. W. F. EDWARDS

Gonville and Caius College,
Cambridge CB2 1TA, UK

1. Sibly, R.M. *Nature* 327, 268 (1987).
2. Fisher, R.A. *The Genetical Theory of Natural Selection*, 2nd edn (Dover, New York, 1958).
3. Turner, J.R.G. *Oxford Surveys in Evolutionary Biology*, 2, 159–96 (Oxford University Press, 1985).
4. Sibly, R.M. & Calow, P. *Physiological Ecology of Animals* (Blackwell, Oxford, 1986).

Investigating the paranormal

SIR—Critics of my analysis of parascientific research¹ provide no new arguments or evidence and the original conclusion of my Commentary — that parascience is pseudoscience — stands. If anything, the structure of their arguments strengthens this conclusion. Couch² suggests that Bell's theorem provides a theory for *psi* phenomena but not a mechanism. Apart from the inconsistency of this argument, it remains doubtful whether Couch (or any-

body else for that matter, including physicists) can actually demonstrate how Bell's theorem helps to explain *psi* phenomena. Couch asks whether millions of people are psychologically abnormal because Waugh's sample of believers had higher neuroticism scores than her sample of disbelievers. Of course not; having a significantly higher score does not imply abnormality but it does point to a trend.

Elitzur³ claims that an empirical observation of personality differences between believers and sceptics is an *ad hominem* argument. If so, then a whole area of psychological investigation on personality differences must be similarly discredited. Elitzur's claim of misrepresentation should perhaps be answered by the two reviewers whose work I previously cited, Akers⁴ and Hyman⁵.

Hyman's selection of experiments was actually provided by Honorton and included all ganzfeld experiments published up to the year 1984. The experiments conducted by Jahn⁶ and Krippner and Ullman⁷ which Elitzur cites as methodologically rigorous are themselves difficult to replicate and Elitzur's enthusiasm for this work may well prove to be misplaced.

Morris⁸ suggests that there is a corpus of paranormal research of which I am unaware which does not contain the flaws outlined in the Commentary, yet he fails to cite it. The reason is that this corpus of evidence does not exist. Such tireless faith in the paranormal hypothesis is hardly rational in the absence of any repeatable paranormal effect.

Stevenson's protestations⁹ may serve some therapeutic function, but like the other critics, he offers no new evidence of a repeatable experiment, only anecdote. Can science be based purely upon anecdote? Like Koestler and Rhine before him, Stevenson assumes that striking coincidences cannot occur by chance, a very serious error for a scientist in his field, as explained in my article. What parapsychologists need to do, if they wish to convince others that they have discovered something, is to produce some coherent theories and some convincing experimental demonstrations. Until then the sceptical majority can hardly be expected to take them seriously.

DAVID F. MARKS

School of Psychology,
Middlesex Polytechnic
Queensway,
Enfield EN3 4SF, UK

1. Marks, D.F. *Nature* 320, 119–224 (1986).
2. Couch, M. *Nature* 321, 465 (1986).
3. Elitzur, A.C. *Nature* 321, 465 (1986).
4. Akers, C. in *Advances in Parapsychological Research* Vol. 4 (ed. Krippner, S.) 112–164 (McFarland, Jefferson, North Carolina, 1984).
5. Hyman, R. J. *Parapsychol.* 49, 3–49 (1985).
6. Jahn, R.G. *Proc. IEEE* 70/2, 136–170 (1982).
7. Krippner, S. & Ullman, M.J. *Am. J. Soc. psychosom. med.* 20, 9–17 (1973).
8. Morris, R.L. *Nature* 322, 680 (1986).
9. Stevenson, I. *Nature* 322, 680 (1986).

Wave functions in the laboratory

A clever measurement of the radiative decay of excited molecular hydrogen has made possible direct estimates of a vibrational wave function, perhaps pointing the way to new tests of old calculations.

PEOPLE calculate wave functions all the time, but can they measure them? The diffraction pattern of an electron beam scattered by a metal may be a good representation of the wave function of a free electron, but that is hardly startling. Now an ingenious piece of spectroscopy of the humble hydrogen molecule by W. Koot, W. J. van der Zande and J. Los (*Phys. Rev. Lett.* **58**, 2746; 1987), from the Dutch Institute for Atomic and Molecular Physics at Amsterdam, seems to have broken new ground in the visualization of a bound-state wave function.

The measurement is that of the spontaneous disintegration of a highly excited electron state of molecular hydrogen which, being a triplet (parallel electron spins), is inaccessible from the singlet ground state, but whose decay leads predominantly to the separation of the two atoms and the emission of a photon. If that were all there is to say, the measurement would be simply that of the frequency of the photon. But molecules have vibrationally excited states as well so that, in a bound-free transition, the vibrational energy is partitioned between the energy of the photon and the kinetic energy of the separating atoms.

Much of the cleverness of the experiment lies in the preparation of hydrogen molecules in a pure state in which the vibrational as well as the electron quantum numbers are known. In the end, Los and his colleagues are able to use a measurement of the energy carried away by the separating hydrogen atoms to calculate the shape of the vibrational wave function. The objective of these sophisticated measurements is not to do the people who make calculations out of a job, of course, but to test the accuracy of the calculations. How?

Los and his colleagues first prepare a mixture of excited states by a charge-exchange reaction with caesium and then use a well-tuned dye laser to excite molecules in one of these (also precisely defined, but metastable) into the state ($g^3\Sigma_g^-$) that generates their bound-free transition. By that transition, the molecules are carried into a state distinguished from the usual ground state of hydrogen only by the parallel alignment of the spins, but which, on that account alone, is an unbound repulsive state.

To obtain the vibrational wave function of the upper state, Los and his colleagues

record the kinetic-energy spectrum of the recoiling hydrogen atoms. (The photon-energy spectrum would serve as well, but covers too wide a range of energy to be detected with a single device.) What they call their translational spectrometer requires that the molecules should be travelling at something like $1,000 \text{ km s}^{-1}$ — the speed of a high-velocity molecular beam — before they are excited by the well-tuned laser and the bond between the two hydrogen atoms is ruptured. The large solid-angle detectors measure the separation both in space and time of the fragments, from which the dynamics of the molecular disintegration can be reconstructed. All this is reminiscent of nuclear disintegrations, which goes to show how far these apparently independent fields have developed their techniques along the same lines.

The spectrum obtained is simply related to the vibrational wave function of the excited molecule by the Franck-Condon principle — the idea that optical transitions are so much more rapid than the motions of atoms in molecular vibrations that the relative positions of the atoms will be unchanged in the transition from the excited molecule to the lower repulsive state. This means that the probability of a transition to a state with a particular separation of the atoms is the probability that the atoms have that same separation in the excited molecule, which is by definition the square of the vibrational wave function. But the kinetic energy of the separate atoms is what they had at the instant of the disintegration plus what they gain as they recoil from each other in the repulsive state. The measured kinetic energy spans the region from 0 to 4 eV, a substantial fraction of the electron excitation of just over 10 eV.

The result is a kinetic-energy spectrum with as many maxima as there are maxima in the square of the vibrational wave function. Because the spectrometer fails to record the fastest atoms, the spectrum turns out to have five maxima rather than the expected six. Although the absolute intensity of the spectrum also depends on the translational wave function of the final repulsive state, that is sufficiently well-known for the authors to be able to calculate the vibrational wave function of the original excited state.

This measurement, likely to be the first of many, may be thought to have some of

its roots in experiments reported nearly a decade ago by J. Tellinghuisen *et al.* (*J. chem. Phys.* **71**, 1283–1291; 1979). There, the objective was to learn something about the interaction between single atoms of sodium and argon, which in their ground states is almost wholly repulsive but which, at very low temperatures, can allow the formation of the curious molecule NaAr which, when formed, can be excited.

Tellinghuisen and his colleagues used measurements of the energy spectrum of transitions from specific excited bound states of the molecule to the usual repulsive state to throw light on the form of the repulsion between ground-state atoms. In doing so, they were paving the way for studies of bound-free transitions such as those now described by Los and his colleagues.

The mere measurement of the vibrational wave function of molecular hydrogen in a particular (and ordinarily inaccessible) state is, naturally, only part of what Los and colleagues are about. Indeed, their main purpose is not to verify what countless calculations have shown, but to estimate the differences between the calculated and measured wave functions so as to estimate the degree to which the excited state is mixed with other states of similar energy.

That is an important goal because it can provide observational tests of calculations of the interatomic forces in the different excited states of molecular hydrogen; the authors promise further measurements of excited molecular hydrogen and arrangements to fill the gap in the high-energy part of the kinetic-energy spectrum.

Even as things are, they find that the difference between the measured and calculated wave functions point to some degree of mixing between states of similar energy, which in turn points to an experimental technique for testing the validity of the Born-Oppenheimer approximation — the cornerstone of simple molecular wave-mechanics in which the motion of electrons and of nuclei is taken to be strictly separable.

The issue is not, of course, whether that principle will be overturned, but whether the domain of its usefulness can be defined by observation. That this can now be considered seriously is a vivid proof of the sophistication of molecular spectroscopy.

Roland Pease

Supernova 1987A

The parent and its environs

Paul Murdin

GREAT interest has been aroused by the supernova SN1987A that occurred in the Large Magellanic Cloud (LMC) this spring. This was partly because of its extreme visibility, being so close; and partly because of the opportunity afforded to study a supernova in more detail than ever before. The neutrino observations are one example of this. The first international conference on the subject* focused on the developing interaction of SN1987A with the surrounding material as well as the nature of the progenitor star.

From astrometry (R. West, European Southern Observatory (ESO); G. White, Anglo-Australian Observatory (AAO)) and the International Ultraviolet Explorer (IUE) observations, it is clear that it was the supergiant Sk-69 202 that exploded. In principle, the star may have been in a multiple system with the other members now masked by the opaque supernova shell (V. Radhakrishnan, Raman Research Institute; R. Gilmozzi, IUE).

Sk-69 202 would have lost a considerable fraction of its original mass (20 solar masses) as it evolved until the explosion occurred (A. Maeder, Geneva Observatory).

Other stars of similar spectral type in the LMC have surface abundances showing helium and nitrogen enhancement, indicating that they have completed hydrogen burning. Archival spectra (W. Wamsteker, IUE) show some evidence (a strong nitrogen line) for this suggestion in Sk-69 202 itself. It is not clear whether a star of this size would remain a blue supergiant, or whether it would become a red supergiant for some time before returning to be a blue supergiant (J. Truran, Max Planck Institute for Astronomy (MPIfA)). The size of the envelope is very sensitive to fine adjustments of the stellar opacity and convection effects (A. Renzini, ESO), which depend on the metal abundance of the parent galaxy. Observation of the number of red supergiants in the LMC (Renzini; Maeder) suggests that Sk-69 202 was probably a red supergiant 10^4 – 10^5 years ago.

Given that the explosion occurred in a relatively small star, the early light curve and rapid cooling of the outer parts of the supernova, as well as its faint luminosity, are explainable (W.D. Arnett, Chicago Univ.) — if the star envelope is dense, then a larger proportion of the core-

collapse energy goes into kinetic energy rather than radiation, and the small envelope expands quickly. The later behaviour of the light curve as it peaked 60 days after outburst is explained (Arnett) by the creation of 0.10 solar masses of radioactive ^{56}Ni in the supernova explosion. Previous type II supernovae have produced less ^{56}Ni (C. Fransson, Stockholm Observatory). In SN 1987A, 93 per cent of the radiated output was in the post-outburst peak, fed by radioactivity (N. Panagia, Space Telescope Science Institute).

The steeply falling density gradient ($\rho \propto r^{-n}$) in the outer envelope of the blue

Lucy; Kirschner) of the mass of material above the photosphere (1 solar mass after 20 days, 15 solar masses after 120 days) are extremely uncertain. It is obviously important to estimate the mass of the supernova and to relate it to the mass of the progenitor star and its loss of mass in the supergiant stages.

It is necessary to understand the history of the star to define the circumstellar environment of the supernova. The progenitor would have been surrounded (Truran; Maeder) by about 10 solar masses of material, generated in the carbon cycle, driven from the surface layers of the star during its supergiant phases. As Sk-69 202 probably passed through a red-supergiant stage with slow loss of mass, followed by a blue-supergiant stage with rapid loss of mass, the circumstellar envelope is in two zones either side of a thin, shocked shell of radius ~ 1 parsec (R.

Chevalier, Virginia Univ.).

Two strong Ca I components in the interstellar absorption spectrum (P. Magain, ESO), moving at 170 km s^{-1} where Ca II is weak, and coming from a gas with a density 10^9 cm^{-3} could lie in this circumstellar shell, outflowing at 100 km s^{-1} from the supernova. Narrow, ultraviolet emission lines that appeared in IUE spectra about 24 May

(Cassatella) in a high-ionization, high-temperature ($T_e \sim 12,000 \text{ K}$), low-density ($n_e \leq 10^5 \text{ cm}^{-3}$), high-nitrogen, low-carbon-abundance region, outside the supernova shock, may be from the nearside of this region (Chevalier), ionized in the ultraviolet outburst and recombining during the next 60 days.

Two radio outbursts followed the birth of SN1987A: the first, which peaked on day 3 (100 millijansky (mJy) at 1 gigahertz (GHz)) and still showing a residual flux of 5 mJy, is probably synchrotron emission from material expanding just outside the supernova shock. This burst started at low radio frequencies. The source was resolved by Australian/South African very long baseline interferometry, with a size of 3 milliarc seconds or more. The upper limit to the emission measure (proportional to the square of the free-electron density) of material outside the region of the supernova emitting in the first burst is $10^6 \text{ cm}^{-6} \text{ parsec}$.

A second radio burst has been reported from São Paulo, 22–25 June, with an intensity of 500 mJy at 22 GHz. This radio burst is not yet observed elsewhere and the flux must be less than 15 mJy at 8 GHz, 5 mJy at 843 MHz and 500 mJy at 86 GHz (R. Manchester, CSIRO; P. Shaver, ESO). The spectrum derived from these values, which is just consistent with the São Paulo data, suggests external free-free absorption, with an emission measure of

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

SN1987A before (left) and after (right, arrow) the explosion.

supergiant progenitor star is preserved during the rapid, supersonic, unturbulent outflow of the supernova, but, according to interpretations of the optical and ultraviolet spectra, the exponent n dropped from about 10 (for the first 5 days) to about 5 (from an age of 5 days to now). This is because the region where the optical depth is 1 (the photosphere) falls deeper into interior regions of the star envelope, as the outer layers expand to become less dense and more transparent. The transparent layers are those which are studied spectroscopically and the unusually rapid development of the optical and ultraviolet spectra was a consequence of the steep density gradient (L. Lucy, ESO), and rapid expansion. It is possible to construct synthetic spectra at more or less normal abundances (D. Branch, Oklahoma Univ.) or with slight helium overabundance (Lucy) that show good agreement with the observational archives of spectra from ESO (R. Fosbury, Space Telescope — European Coordinating Facility), South African Astronomical Observatory (J. Menzies) and IUE (A. Cassatella, IUE; R. Kirschner, Smithsonian Center for Astrophysics). Even the puzzlingly strong barium line at $6,050 \text{ Å}$ (Fosbury; J. Dachs, Bochum Univ.), rarely seen in supernovae, is actually from a normal barium abundance. The synthetic spectra are difficult to calculate in detail and estimates (P. Höflich, MPIfA;

* European Southern Observatory Workshop on Supernova 1987A, Garching-bei-München, 6–8 July 1987.

$5 \times 10^8 \text{ cm}^{-6}$ per sec. For the two bursts to be consistent, the absorbing material must be fragmented or asymmetrical. The second burst started with high frequencies, suggesting that the source may be a pulsar (F. Pacini, Arcetri Observatory), although dispersion would make the pulses unresolvable. The characteristic radio emission seen in other type II supernovae is not expected for 10–50 years, when the supernova shock overtakes the wind-shocked shell.

The circumstellar material from the mass-loss phase of the red supergiant, outside the wind-shocked shell, might be expected to be dusty and, when warmed by the supernova outburst, would be expected to echo infrared emission. The Kuiper Observatory, airborne on 21 and 29 April, detected (S. Drapatz, Max Planck Institute for Experimental Physics) not only the supernova photosphere, but also an infrared excess between 7.8 and $10 \mu\text{m}$ with a temperature of 300 K and a flux of $2 \times 10^{38} \text{ erg s}^{-1}$, attributable (Chevalier; Drapatz) to this echo.

Narrow ultraviolet lines that appeared in May could arise from inside the wind-shocked shell (Kirschner), as could a narrow zero-velocity Ca II feature that appeared in optical spectra on 15 June (Fosbury), and interstellar Al III which changes its equivalent width as measured by IUE (M. Grewing, Tübingen Univ.).

From 15 May, the spectrum at infrared wavelengths greater than $5 \mu\text{m}$ showed (A. Moneti, ESO) a second, increasing component of flux (black-body temperature = 1,240 K). Silicate and graphite dust will not have condensed from the supernova ejecta yet, and the temperature is too high for this second component to be the infrared echo. Free-free emission

from the blue-supergiant wind could be the source. One-dimensional infrared speckle observations (A. Chalabaev, Haute Provence) at $3.6 \mu\text{m}$ (but not 2.2 or $5 \mu\text{m}$) show the supernova to be resolved in east–west and north–south scans to between 50 and 200 milliarc seconds (0.02 parsec) but the ratio of the flux of the resolved component to the supernova background is uncertain.

Optical speckle interferometry of the supernova (P. Meikle, Imperial College London; Kirschner) shows an object nearby (christened the mystery spot after a California amusement park). The mystery spot lies 65 milliarc seconds south of the supernova. It has been detected in three red wavebands with a separation from the supernova on days 30–38 of 59 milliarc seconds and on day 50 of 74. (This is marginal evidence for outward motion of the spot at 0.46 times the speed of light; but the same data could show that it is stationary at a mean distance of 18 light days from SN1987A.)

About 3 magnitudes fainter than the supernova, the mystery spot seems to have a red continuum spectrum; it is 100 times brighter than any other star in the LMC and, astronomers are sure, was not there before the outburst. Although it subtends a solid angle of only about 2 per cent of a sphere centred on the supernova, it radiates at 10 per cent of the radiated power of the supernova ($L \sim 5 \times 10^{38} \text{ erg s}^{-1}$). One explanation for it was that a hole punched in the supernova shell by a second star in the Sk-69 202 system focused a relativistic jet on to nearby material. □

Paul Murdin is at the Royal Greenwich Observatory, Hailsham BN27 1RP, UK.



Giraffes in their 'anti-gravity suits'.

free-ranging animals, and have additionally recorded osmotic pressures and blood flow in sedated animals standing still. The means of the data in the feet of standing animals show $P_a \sim 260 \text{ mm Hg}$; $P_v \sim 150 \text{ mm Hg}$; and $P_t \sim 44 \text{ mm Hg}$. In moving animals, however P_a varies between 70 and 380 mm Hg; P_t between –120 and +80 mm Hg; and P_v varies hugely, between –250 and +240 mm Hg.

Features of the peripheral circulation that inhibit oedema in a standing giraffe are first, a high resistance to flow in the thick-walled arterioles, which keeps P_a and hence capillary pressure P_c well below P_v ; and second, a very tight skin and fascia in the lower legs (an 'anti-gravity suit'), which allows P_t to be much higher than in man (where $P_t \sim 0$). Even so, there is a net filtration pressure of more than 80 mm Hg, and quietly standing giraffes will be susceptible to some oedema although one might expect P_t to continue to rise as oedema proceeds, assuming the fascia to be effectively inextensible.

In the ambulant giraffe, however, the well-known 'muscle pump' comes into play, as in man, squeezing blood up out of the lower veins as the skeletal muscles contract, and sucking it in again through the capillaries as they relax, backflow in the veins being prevented by the valves. A high arteriolar resistance would then explain the very low observed values of P_v (and hence P_c), because it prevents the compliant veins from filling rapidly; this explanation can be checked from the phase relations between P_a and P_v if the radiotelemetry equipment allows them to be recorded as functions of time.

The siphon model¹ of the circulation in the head and neck is based on the fact that if the blood vessels are rigid, or at least stiffly distended, gravitational pressure changes are irrelevant, $P_a - P_v$ at the heart being determined purely by a constant

Haemodynamics

How giraffes prevent oedema

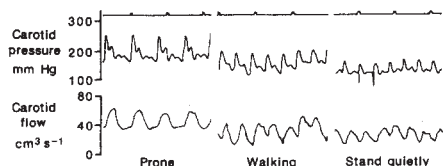
T.J. Pedley

AN upright giraffe, by analogy with humans, ought to suffer massive oedema in its feet; moreover, when it lowers its head to drink, the blood should rush down into it and be unable to flow up again. New pressure measurements reported on page 59 of this issue¹ by Hargens *et al.* show why neither of these things happens, and also contain some surprising observations of highly variable venous pressure (P_v) in the leg and of a counter-gravitational gradient of P_v in the neck, which greatly improve our understanding of the circulation of blood in the giraffe. The latter result, in particular, demonstrates conclusively that the jugular vein is normally highly collapsed and therefore the circulation in the head and neck does not resemble a siphon, despite recent opinion

to the contrary².

The gravitational (or hydrostatic) gradient of pressure with height means that, in an upright animal of height 3.5 m, the mean blood pressure in an artery in the head is as much as 110 mm Hg (about 1.5 m H₂O or 15 kPa) lower than at the level of the heart, which is about 200 mm Hg above atmospheric, double the human value, to stop the carotid artery from collapsing. In the foot, mean arterial pressure (P_a) will be 110 mm Hg greater, that is, 310 mm Hg, which, if the peripheral circulation were the same as in man, would lead to pooling of blood in the veins and to oedema.

Hargens *et al.*¹ made radiotelemetry measurements of P_a , P_v and tissue pressure (P_t) (all relative to atmosphere) in



Phasic waveforms of carotid pressure and flow in the giraffe during spontaneous activity (redrawn Fig. 6 from ref. 5).

viscous resistance. But blood vessels are compliant and can collapse flat to a very small cross-sectional area when the internal pressure falls significantly below the external, and even in man the jugular vein is often in a collapsed state⁴. (Veins in the head do not collapse because the skull is a constant-volume box filled with incompressible tissue and fluid, so incipient collapse causes a corresponding fall in extra-vascular pressure.) The gravitational contribution to P , at the top of a standing giraffe's neck is around -100 mm Hg (right atrial pressure being close to atmospheric), not the observed $+13$ mm Hg (ref. 1). The difference cannot be accounted for by compartmentalization of the vein, as although there are valves in the jugular vein of a giraffe (unlike man), they work the wrong way round. The required mean pressure gradient of about 80 mm Hg m^{-1} must come from viscous resistance to flow in the vein itself. However, if the jugular vein were fully open, with diameter 2.5 cm (ref. 2), it would take a flow rate of above $1,500$ $l\ min^{-1}$, 500 times the flow in the carotid artery⁵, to generate this gradient, according to Poiseuille's law. Realistic flow rates

require a much larger resistance which can come only from severe collapse of the vein. The analogy is closer to a waterfall than to a siphon⁶.

What about venous return with the head down? This is where the valves come in, but as in the legs it requires a muscle pump to help move the blood up, and this I believe comes from rhythmic movements of the jaws.

All the data I have discussed above refer to mean blood flow, averaged over many heart beats. But, nearly 20 years ago, Van Citters *et al.*⁵ measured the shapes of the pressure and flow-rate pulses in the carotid arteries of free giraffes, and found dramatic differences between prone, quietly standing and walking animals (see figure). I have seen no published explanation of these differences, although it should not be too hard to find one as the giraffe carotid artery is long and straight, so that relatively simple fluid-mechanical theories⁴ should be applicable. □

1. Hargens, A.R., Millard, R.W., Pettersson, K. & Johansen, K. *Nature* **329**, 59–60 (1987).
2. Badeer, H.S. *Comp. Biochem. Physiol.* **83A**, 207–211 (1986).
3. Burton, A.C. in *Physiology and Biophysics* (eds Ruch, T.C. & Patton, H.D.) 523–542 (Saunders, Philadelphia, 1965).
4. Caro, C.G., Pedley, T.J., Schroter, R.C. & Seed, W.A. *The Mechanics of the Circulation* (Cambridge University Press, 1978).
5. Van Citters, R.L., Kemper, W.S. & Franklin, D.L. *Comp. Biochem. Physiol.* **24**, 1035–1042 (1968).
6. Permutt, S., Bromberger-Barnea, B. & Bane, H.N. *J. appl. Physiol.* **18**, 924–932 (1963).

T.J. Pedley is in the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 9EW, UK.

Organic conductors

Inclusion of chalcogens raises electron mobility

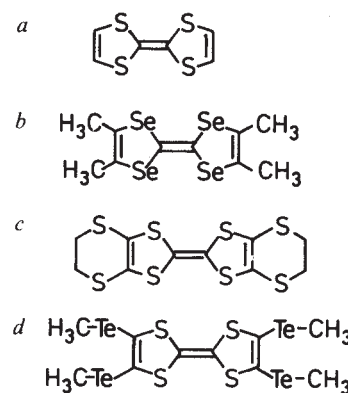
Richard Friend

SEMICONDUCTORS, metals and superconductors are all represented in the class of organic molecular solids with delocalized π -electron systems, that have been investigated for two decades. Milestones in this field have been the discovery of good metallic conductivities in TTF-TCNQ (ref. 1), superconductivity near 1.2 K in TMTSF single-stack salts such as $(TMTSF)_2ClO_4$ (ref. 2) and up to 7 K in $(BEDT-TTF)_2I_3$ (ref. 3). Common to these examples is the presence of sulphur or selenium within the donor molecules (*a–c* in the figure). The beneficial effects of the chalcogens (group-4 elements) have long been recognized, and many have tried to build them into other donor molecules. The inclusion of tellurium into such systems has not been easy, but, as they report on page 39 of this issue, Inokuchi *et al.*⁴ have synthesized the molecule $TTeC_1$ —

TTF (*d* in the figure), and find that it has a relatively high electron mobility.

The composition and structure of these materials conform to a well-established pattern. The organic metals and superconductors are always systems with a non-integral oxidation state for the donor (and/or the acceptor) molecule. This gives partial filling of an electron band, which is a prerequisite for metallic conduction. The molecules are planar and arranged to form one-dimensional stacks; overlap of the π -electron orbitals on adjacent molecules within the stack then allows delocalization of charge along these stacks.

Contact of the π -electrons between the chains is often poor, and these materials can be extremely anisotropic, or quasi-one-dimensional, conductors. The chalcogens provide stronger coupling of electrons between adjacent stacks, making



Four organic donor molecules, *a*, TTF; *b*, TMTSF; *c*, BEDT-TTF; *d*, $TTeC_1$ -TTF. Each contains group-4 elements (chalcogens), sulphur, selenium and tellurium. Is a new effect heralded by the new compound (*d*) synthesized by Inokuchi *et al.*?

the charge transport more three-dimensional, thus their inclusion improves the conducting properties of organic conductors.

Metals are inherently unstable in one dimension, as was first recognized by Peierls⁵. It is always energetically favourable to distort the structure so that there is an energy gap between the highest-occupied and lowest-unoccupied electron energy levels, achieved by setting up either a superlattice by displacement of the lattice (periodic lattice distortion coupled to a charge-density wave) or through a modulation of the electron-spin density (spin-density wave). The directionality of the π -electron systems in the known organic conductors usually gives quasi-one-dimensional properties, with anisotropy ratios of 100 or more for the conductivity parallel and perpendicular to the stacks.

Among the hundreds of charge-transfer salts that have been investigated, most show metallic properties at room temperature, but pass through a phase transition to one of the superlattice phases to become semiconducting as the temperature is lowered. The study of the superstructures that are obtained at low temperatures is itself interesting, but the study of the high-conductivity metallic phases is considerably hampered if the metallic phase cannot be preserved to low temperatures.

TMTSF, with four saturated methyl groups on the outside of the molecule, was originally investigated as a variant of TTF likely to give it a more strongly one-dimensional behaviour than TTF itself. The salts of composition $(TMTSF)_2X$ (where X is a monovalent anion) were found², however, to behave in the opposite way. Many of these salts remain metallic to liquid-helium temperatures (although high pressures are sometimes required), and several of them become

superconducting near 1.2 K. Inspection of the selenium-selenium contacts between adjacent stacks shows that these are indeed strong in at least one of the inter-stack directions. The measured anisotropy in conductivity between this direction and the stack axis is as low as 20, and these materials are then anisotropic, but two-dimensional, metals.

To the solid-state physicist, the acid test for high-mobility, metallic behaviour is the measurement of oscillations in the resistivity (de Haas-Shubnikov oscillations) or magnetic susceptibility (de Haas-van Alphen oscillations) with a magnetic field applied. These arise from quantization of the delocalized conduction electrons by the field into discrete levels (Landau levels). The period of these oscillations measures directly the cross-sectional area in momentum space, normal to the applied field, of the 'pockets' of electrons or holes at the Fermi energy. This gives very direct information about the dynamics of the conduction electrons.

These phenomena are only observed if the electrons can make 'cyclotron' orbits without being scattered by lattice vibrations or impurities, and it is necessary to use high-purity samples at liquid-helium temperatures to minimize these scattering processes. Several measurements of these magnetic oscillations in TMTSF salts^{6,7}, demonstrate unequivocally that coherent electron motion can be achieved between the TMTSF stacks.

Salts of BEDT-TTF show even stronger interstack interactions than those formed with TMTSF, and the conductivity anisotropy is often very small. In both classes of salts, however, the carbon-based π -electrons still dominate intrastack delocalization. In the tellurium-containing molecule of Inokuchi *et al.* reported in this issue², it seems that this may no longer be true. As Inokuchi *et al.* point out, the network formed by the tellurium atoms alone is probably sufficient to allow the easy transport of charge. It is then a matter of semantics as to whether such materials are to be called organic conductors or tellurium compounds. However this question is resolved, there is interesting science in store, and the charge-transfer salts formed with TTeC₁-TTF are eagerly awaited. □

1. Coleman, L.B. *et al.* *Solid St. Commun.* **12**, 1125 (1973).
2. Bechgaard, K., Carneiro, K., Olsen, M., Rasmussen, F.B. & Jacobsen, C.S. *Phys. Rev. Lett.* **46**, 852 (1981).
3. Baram, G.O. *et al.* *J. exp. theor. Phys. Lett.* **44**, 376 (1987).
4. Inokuchi, H. *et al.* *Nature* **329**, 39–40 (1987).
5. Peierls, R.E. *Quantum Theory of Solids* (Oxford University Press, 1955).
6. Bouffard, S., Ribault, M., Jérôme, D. & Bechgaard, K. *J. Phys. Lett.* **44**, L285 (1983).
7. Schwenk, H., Parkin, S.S.P., Schumaker, R., Greene, R.L. & Schweitzer, D. *Phys. Rev. Lett.* **56**, 667 (1986).

Richard Friend is in the Cavendish Laboratory, Cambridge University, Madingley Road, Cambridge CB3 0HE, UK.

Neurobiology

Molecules underlying memory

Mary B. Kennedy

SHORT-TERM information storage in the nervous systems of both vertebrates and invertebrates often involves transient covalent modifications of proteins that control the strength of synapses¹. In contrast, long-term information storage, such as that involved in the formation of memories, probably involves changes in gene expression that ultimately alter the structure of the synapse^{2–4}. There is much debate about the extent of overlap between the mechanisms underlying short-term and long-term information storage^{2,3}. The report by Greenberg *et al.* on page 62 of this issue⁵ provides a possible molecular link between these two processes in the synapses mediating the familiar siphon withdrawal reflex in the marine mollusc *Aplysia*. Training paradigms that produce long-term sensitization of the reflex cause a small but significant reduction of the ratio of regulatory subunits of the cyclic (c)AMP-dependent protein kinase to catalytic subunits in sensory neurons. This reduction could produce a substantial increase in free, active catalytic subunits and contribute to long-term sensitization.

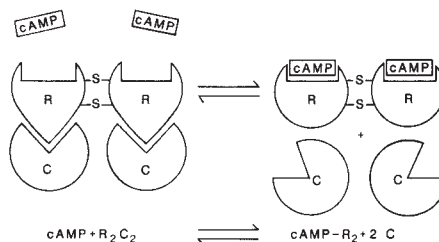
In response to a threatening stimulus, *Aplysia* withdraws its siphon towards its mantle shelf. A brief train of electrical shocks to the tail sensitizes this response so that for about an hour the animal withdraws its siphon more vigorously in response to further stimuli. The sensitization seems to result from the release of neuromodulators onto sensory synapses that contact the siphon motor neurons². These modulators increase the synthesis of the second-messenger cAMP within the synapses, activating the enzyme cAMP-dependent protein kinase. Protein kinases are central elements in the regulatory networks of all cells. They catalyse the transfer of phosphate from the universal cellular energy reservoir ATP to the side chains of serine, threonine or tyrosine residues within proteins. In so doing they often alter the functions of the recipient proteins. Within the sensory synapse, the activated kinase phosphorylates a protein which then closes a crucial potassium channel (the S-channel) in the presynaptic membrane. The effect of this closure is to retard the repolarization of the membrane each time the sensory synapse is activated and thus to allow more calcium to flow into the synapse through voltage-dependent calcium channels. Ultimately, this causes the release of more transmitter. The behavioural effect is enhanced contraction of the siphon (sensitization).

Short-term sensitization lasts as long

as the concentration of cAMP is elevated and the kinase is activated (about one hour for each sensitizing shock)². On the other hand, a series of closely spaced tail shocks delivered over a few hours each day produces long-term sensitization which can persist for days to weeks, depending on the number of shocks and the number of days of training⁶. What are the molecular mechanisms that account for this persistence? Experiments with *Aplysia* show that protein synthesis is required for long-term but not for short-term sensitization⁷. Nevertheless, long- and short-term sensitization involve changes in the same synapses and in the same potassium channel.

In 1982, Kandel and Schwartz² suggested a simple mechanism to explain the protein-synthesis requirement. The cAMP-dependent protein kinase is a tetramer of two regulatory subunits and two catalytic subunits (see figure). When the cAMP concentration is low, the regulatory subunits are bound to the catalytic subunits and inhibit them; when the concentration is elevated, the regulatory subunits bind cAMP which causes a change in conformation and 'release' of the active catalytic subunits (see figure). Kandel and Schwartz suggested that persistent increases in cAMP produced by repeated sensitizing stimuli might induce the expression of a new isozyme of the regulatory subunit with a higher affinity for cAMP. They imagined that this could ultimately produce tonic activation of a portion of the catalytic subunits at resting cAMP levels. They produced no evidence to support that hypothesis, but the report in this issue⁵ presents a new twist.

The generation of long-term sensitization causes a small but significant reduction in regulatory subunits in sensory neurons.



The cAMP-dependent protein kinase consists of two cAMP-binding regulatory subunits (R) and two catalytic subunits (C). At low cAMP concentrations the regulatory subunits bind to and inhibit the catalytic subunits. At higher concentrations the regulatory subunits bind cAMP and release the catalytic subunits which can then phosphorylate protein substrates. S-S, disulphide bonds.

The reduction persists for at least 24 hours after the training stimuli. It is not seen in sensory neurons from animals that have been stimulated to produce only short-term sensitization. Greenberg *et al.*⁵ hypothesize that at resting cAMP levels the observed reduction in regulatory subunits shifts the equilibrium of the kinase towards free catalytic subunits, resulting in a persistent closure of potassium channels and persistent sensitization. The authors do not demonstrate that new protein synthesis is required for the reduction of regulatory subunits, but suggest that new synthesis of a protease or a repressor of the regulatory subunit gene underlie long-lasting reductions in the regulatory subunit. Earlier studies of mammalian cells show that free regulatory subunits are highly susceptible to proteolysis⁶, and that concentrations of the regulatory and catalytic subunits can be independently regulated⁸.

This mechanism adds to a growing list of long-lasting biochemical changes in

neurons that can be produced by transient changes in the environment^{4,9}. A difficult task that lies ahead is to understand how these mechanisms interact within the intricate and interdependent regulatory networks of the neuronal cytoplasm; which of the mechanisms are important for the physiology of different synapses, and which are in fact important in the encoding of memories. □

1. Kaczmarek, L.K. & Levitan, I.B. (eds) *Neuromodulation: The Biochemical Control of Neuronal Excitability* (Oxford University Press, 1987).
2. Kandel, E.R. & Schwartz, J.H. *Science* **218**, 433–443 (1982).
3. Golet, P. *et al.* *Nature* **322**, 419–422 (1986).
4. Black, I.B. *et al.* *Science* **236**, 1263–1268 (1987).
5. Greenberg, S.M., Castellucci, V.F., Bayley, H. & Schwartz, J.H. *Nature* **329**, 62–64 (1987).
6. Frost, W.N., Castellucci, V.F., Hawkins, R.D. & Kandel, E.R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8266–8269 (1985).
7. Steinberg, R.A. & Agard, D.A. *J. biol. Chem.* **256**, 10731–10734 (1981).
8. Richards, J.S. & Rolfes, A.I. *J. biol. Chem.* **255**, 5481–5489 (1980).
9. Schwartz, J.H. & Greenberg, S.M. *A. Rev. Neurosci.* **10**, 459–476 (1987).

Mary B. Kennedy is in the Department of Biology at the California Institute of Technology, Pasadena, California 91125, USA.

Atomic physics

Bragg scattering revisited

Juha Javanainen

WAVE-PARTICLE duality, as described by the de Broglie formula, was first confirmed in experiments by G. P. Thomson, and by C. J. Davison and L. H. Germer in the 1920s in which beams of electrons were diffracted (wave-like) by crystals. These early substantiations of quantum theory were the ancestors of modern electron- and neutron-diffraction techniques. But the ambiguity can be pushed further. Material particles can be diffracted by light-beams, a possibility first discussed by P. L. Kapitza and P. A. M. Dirac 50 years ago (*Proc. Camb. phil. Soc. math. phys. Sci.* **29**, 297–300; 1933). But it is only now that the effect has been demonstrated unequivocally and in full agreement with theory by David Pritchard and co-workers in MIT (Martin, P. *et al.* *Phys. Rev. A* **36**, 2495–2498; 1987).

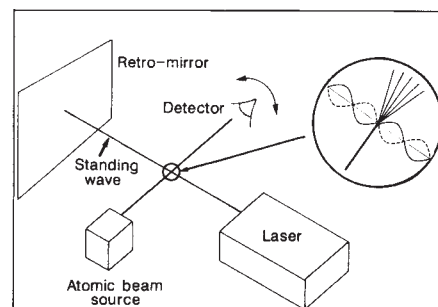
In their experiments a tightly focused, monoenergetic atomic beam is shot across a standing light wave at an accurately controlled angle close to 90°, and the transverse profile of the atomic beam is recorded downstream far from the interaction region. As the period of the standing wave is half the wavelength of light $\lambda/2$, and the de Broglie wavelength of the atoms with the momentum p is h/p (where h is Planck's constant), the Bragg diffraction law predicts atom scattering at multiples of the angle $2h/\lambda p$. Indeed, the experiments show a splitting of the atomic beam into components whose transverse momenta differ by integer multiples of $2h/\lambda$.

Thanks to the enormous resonance

enhancement of the interaction strength when the laser frequency is close to an atomic transition frequency, the Kapitza–Dirac effect is easier to observe with atoms than with electrons that were originally suggested. The trick that made the experiment of Martin *et al.* work, though, was to avoid tuning the laser too close to the resonance. The probability that the atom is excited, and hence the probability of a spontaneous emission, then remains small. In the experiments the coherence of the light–atom interaction is preserved uninterrupted by spontaneous emissions, even though the time of flight of the atoms through the laser beams is several spontaneous lifetimes. As an additional windfall with this detuning, the theory becomes very simple. For instance, the unavoidable time variation of the light intensity seen by the atom while it crosses the laser beam can be taken into account exactly.

Nonetheless, in earlier experiments (Gould, P. L., Ruff, A., Pritchard, D. E., *Phys. Rev. Lett.* **56**, 827–830; 1986) the spreading of the atomic beam would agree only with that predicted for a laser intensity half that measured. To improve both the experimental techniques and the theory, the authors varied with high accuracy the angle between the atomic beam and the laser beam around the precisely orthogonal alignment, and included the effect of alignment in the theory.

The new results show that the entrance angle critically affects the Kapitza–Dirac



Schematic diagram of the layout of Martin *et al.*'s experiment. The atomic sodium beam intersects the standing laser wave at right angles and is diffracted.

scattering. If the atoms do not impinge precisely perpendicularly to the laser beam, they have, of course, a component of velocity parallel to it. If the atoms move more than half a wavelength along the standing wave before leaving the interaction region, the periodic light lattice effectively becomes averaged out and diffraction diminishes. A milliradian misalignment could have caused the earlier discrepancy — in the new experiments the agreement with theory is impressive.

Another way of looking at the experiment is that the standing light wave, consisting of two opposed travelling waves, contains photons whose momenta along the direction of the laser beam are $\pm h/\lambda$. Suppose the atom first absorbs a photon with momentum, say $+h/\lambda$, and in so doing acquires a recoil kick in the plus direction. If the next optical process is induced emission by a photon with momentum $-h/\lambda$, the recoil kick on the atom will again be in the plus direction. The atom thus returns to its ground state

ALSO reversing some usual notions of light and matter, V. I. Balykin *et al.* have constructed a 'mirror' made of light that reflects a beam of sodium atoms (*Pis'ma éksp. teor. Fiz.* **45**, 282–284; 1987; to appear in translation this week in *J. exp. theor. Phys. Lett.* **45**, 353–356; 1987). When light is totally internally reflected off a quartz/vacuum interface, an evanescent field penetrates about half a wavelength into the vacuum. It is this field that makes the 'mirror'.

With a grazing incidence, nearly all the sodium beam is reflected off the 'mirror'. As the angle of incidence is increased (to $\sim 20^\circ$ to the plane of the quartz) the reflectance reduces to about 10 per cent, because most of the atoms reach the quartz face itself and are scattered diffusely — a problem that can be overcome by increasing the laser power. Balykin *et al.* look forward to constructing a concave mirror that could focus an atomic beam down to its de Broglie wavelength, opening the way to new studies of atomic optics.

R. P.

with an accumulated net change of momentum $2h/\lambda$.

This alternative description of the Kapitza–Dirac effect explains why the momentum changes by multiples of twice the photon momentum, and illustrates the multifacet nature of quantum mechanics in that it invokes the particle picture for both the light and the atoms. More importantly, however, it also puts the Kapitza–Dirac effect in the same family with the applications of the photon recoil such as laser cooling of atoms and trapped ions, and optical atom traps, all currently at the most active forefront of optical physics. Until very recently, light pressure was a field of many theories and speculations

and few experiments. The situation is changing rapidly but, as it comes to quantitative comparisons between theory and experiment, the work of the MIT group may stand unsurpassed for some time to come. It also demonstrates for the first time fully coherent light pressure action of a standing wave. Realizations of such unique quantum-mechanical effects like band structure of the translational energy of atoms in a standing wave ‘crystal’ of light analogous to Bloch states of electrons may eventually open up, with unforeseen applications. □

Juha Javanainen is at the Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA.

Mathematics

The symplectic camel

Ian Stewart

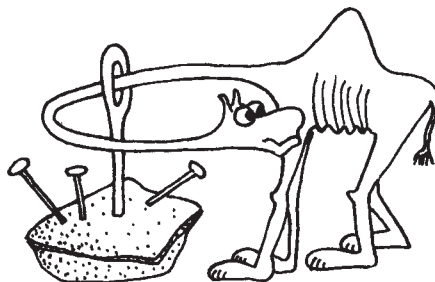
THE motion of matter is one of the richest sources of mathematical concepts. It is possible to trace a continuous thread from the experiments of Galileo Galilei and the empirical laws of Johannes Kepler, by way of Isaac Newton, Joseph-Louis Lagrange, and the optical/mechanical analogies of William Rowan Hamilton, to a large part of the mainstream of present-day mathematics. Mechanics leads to differential equations, and thence to manifolds, Lie groups and measure theory. The normal modes of an oscillator lead to quadratic forms and linear algebra. Heat flow and wave motion lead to Fourier series and functional analysis. But the concept that potentially is the most far reaching of all is currently familiar only to mathematicians and theoretical physicists. It is inspired by a geometrical interpretation of Hamilton’s general formalism for mechanics, an abstruse type of geometry referred to as ‘symplectic’.

The importance of symplectic geometry for mechanics is now clear, thanks in particular to the efforts of the Russian and American schools of dynamical systems over the past 30 years. But a broader movement is now gathering momentum. Vladimir Arnold of Moscow State University has recently laid down what might be termed the manifesto of symplectic mathematics (*Russian Math. Surv.* **41**, 1986). He adduces impressive evidence that a new branch of mathematics, symplectic topology, is coming into bloom. In his book *Catastrophe Theory* (Springer, Berlin, 1987) he goes further, saying “As each skylark must display its comb, so every branch of mathematics must finally display symplectization”. By this he means that many mathematical concepts have analogies within the world of symplectic geometry.

The symplectic saga is a demanding

one, representing a high level of mathematical abstraction whose first steps were taken over two centuries ago when Lagrange introduced generalized coordinates. But it is a story worth telling, in which fundamental properties of the natural world interact with deep and often counterintuitive mathematics. This is not abstraction for its own sake, but as a necessary step to reveal the conceptual essence of physical phenomena: it is the applied mathematics of the twenty-first century in the making.

The word symplectic was coined by Hermann Weyl in his famous treatise *The Classical Groups* (Princeton University Press, 1939), which is about the groups of



Cosgrove

rigid motions of various basic kinds of multidimensional geometry. For ordinary euclidean geometry, the rigid motions form the orthogonal group. If we seek a unified viewpoint, other groups also demand attention. Weyl devoted very little space to the symplectic group — it was then a rather baffling oddity which presumably existed for some purpose, though it was not clear what. Now we know: the purpose is dynamics.

In ordinary euclidean geometry the central concept is distance. To capture the notion of distance algebraically we use the inner (or scalar) product $\mathbf{x} \cdot \mathbf{y}$ of two vectors

$\mathbf{x}$ and $\mathbf{y}$. If $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ are vectors in the plane, then $\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2$; a similar formula holds in higher dimensions. All the basic concepts of euclidean geometry can be obtained from the inner product. In particular a transformation T is a rigid motion if, and only if, it preserves the inner product, $T\mathbf{x} \cdot T\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.

The inner product is a bilinear form — the terms look like $x_i y_j$. Replacing it by other bilinear forms creates new kinds of geometry. Symplectic geometry corresponds to the form $x_1 y_2 - x_2 y_1$, which is the area of the parallelogram formed by the vectors $\mathbf{x}$ and $\mathbf{y}$. Note the minus sign: it leaves footprints all over the symplectic landscape. This symplectic form provides the plane with a new kind of geometry, in which every vector has length zero and is at right angles to itself. There are analogues in spaces of any even number of dimensions.

Can such bizarre geometries be of practical relevance? Indeed they can: they are the geometries of classical mechanics. In Hamilton’s formalism, mechanical systems are described by the position coordinates $q_1, \dots, q_n$, momentum coordinates $p_1, \dots, p_n$, and a function H of these coordinates (nowadays called the hamiltonian) which can be thought of as the total energy. Newton’s equations of motion take the elegant form $dq_i/dt = \partial H/\partial p_i$, $dp_i/dt = -\partial H/\partial q_i$. When solving Hamilton’s equations it is often useful to change coordinates. But if the position coordinates are transformed in some way, then the corresponding momenta must be transformed consistently. Pursuing this idea, it turns out that such transformations have to be the symplectic analogues of rigid euclidean motions. The natural coordinate changes in dynamics are symplectic. This is a consequence of the asymmetry in Hamilton’s equations, whereby dq/dt is plus $\partial H/\partial p$, but dp/dt is minus $\partial H/\partial q$ — that minus sign again.

So far, I have been talking of symplectic geometry, the ‘rigid motions’ of the symplectic world. For symplectic topology we must be more flexible, and use transformations which ‘in the small’ look like symplectic rigid motions. In the jargon these are called symplectomorphisms, but I shall refer to them as symplectic mappings. In the plane the symplectic form represents area, so a symplectic rigid motion is a linear transformation that preserves area. To add flexibility, we relax the condition of linearity. A symplectic mapping of the plane is thus any transformation that preserves area. But shapes can change drastically. For a mental picture, think of the plane as an incompressible fluid and a symplectic mapping as something that stirs the fluid around. (This is not just a picture: fluid mechanics can fruitfully be recast in symplectic language.)

Differential topology is the study of

smooth mappings of a manifold. Analogously, symplectic topology is the study of symplectic mappings of a symplectic manifold. Arnold lays the programme down in these words: "Questions of symplectic topology can be considered as questions of ordinary topology in the presence of additional structure. But of much greater interest to me is not the use of ordinary topology in the study of objects of symplectic geometry, but divining symplectic results by means of 'symplectization'. Symplectization transforms not only the initial objects (manifolds, maps...) but also the whole theory. For example the concepts of boundaries and homology theories in symplectic topology are different things from the ordinary ones. The dimension of a 'symplectic boundary' should be not one, but two less than the dimension of the original manifold."

The earliest theorem of symplectic topology, invented long before such a subject existed, is Henri Poincaré's "last geometric theorem", arising from a problem in celestial mechanics. The theorem states that an area-preserving (symplectic) transformation of an annulus (region between two circles), which moves the two boundary circles in opposite directions, has at least two fixed points. Such fixed-point theorems are very powerful: this theorem, proved by George Birkhoff in 1913, implies the existence of periodic orbits in the motion of three bodies under gravity. If the transformation is not symplectic, there need not be any fixed points; so symplectic topology has its own distinctive character.

Much of this character is as yet very hazy: mathematicians are only beginning to learn how to think symplectically. To illustrate this point, consider Arnold's question "Can a symplectic camel go through the eye of a needle?" In higher dimensions, as in the plane, symplectic mappings preserve volume. But does being symplectic impose more restrictions than this? Clearly a volume-preserving camel can pass through an arbitrarily small aperture: just stretch the camel out until it is very long and thin, and thread it through. In contrast, Mikhael Gromov has proved (unpublished work) that a symplectic camel cannot pass through, because its ribs get in the way. The camel is suitably mathematical: a sphere. Its ribs are inequalities. None of this is obvious or — yet — intuitive.

Arnold discusses a wide variety of solved and unsolved questions in symplectic topology. A large bulk comprises extensions and generalizations of Poincaré's geometrical theorem, for example the recent proof by C.C. Conley and E. Zehnder (*Commun. pure appl. Math.* 37, 207–253; 1984) of a multi-dimensional analogue conjectured by

Arnold. Others concern symplectic analogues of standard topological ideas, such as knots and homology groups.

One application is to optics. When light rays pass through an optical system they may form caustics — bright curves and surfaces where the light comes to a focus. A list of caustic morphologies was proposed, based on singularity theory; it also applies to galaxy formation (see my News and Views article in *Nature* 323, 397; 1986). According to the list, the commonest morphological change is the 'flying pancake', in which a hollow lens of dense material is formed. It is the three-dimensional analogue of a dressmaker's tuck. The list remained unchallenged for some time, until John Nye and John Hannay noticed that the flying pancake cannot occur as a morphology of an optical caustic (*Optica Acta* 31, 115–130; 1984), although it can occur in galaxy formation. The underlying reason is now understood: the evolution of galaxies is governed by ordinary topology, but that

of caustics is governed by symplectic topology. This imposes extra constraints and prunes the list of morphologies for caustics. When applying a general mathematical philosophy, it is crucial to work in the 'right' context.

Anyone reading Arnold's paper will be struck by the absolutely vast range of ideas and problems opened up by the process of symplectization. Mathematicians must have felt much this way when they first realized that complex numbers were more than just one extra gimmick: virtually every idea in mathematics, from the geometry of curves to the analysis of partial differential equations, was ripe for complexification. Mathematics exploded overnight. A very large part of mathematics is now waiting to have the secret symplectic side to its nature brought out of the closet. We are witnessing just the tip of a symplectic iceberg. □

Ian Stewart is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

Seasonal depression

Ironies of animal modelling

John D. Hallonquist and N. Mrosovsky

IN the new revision¹ of a manual of mental disorders widely used by psychiatrists, recognition is given to a new variant of affective (mood) disorders: that with a seasonal pattern. At a recent workshop*, investigators from North America and Europe met to discuss 'seasonal affective disorder', its clinical features and epidemiology, the usefulness of animal models, and the efficacy of phototherapy.

One type of seasonal mood disorder is the regular recurrence of depression in winter, in the absence of any obvious seasonally related psychosocial stressors. Even if this depression is left untreated, it commonly remits in the spring. It occurs more often in women, and onset in early adulthood is common².

Together with mood changes, seasonal affective disorder is characterized by lethargy, loss of libido and frequently, in contrast to most cases of non-seasonal major depression, by oversleeping, overeating, craving for carbohydrates and weight gain. Because changes in activity and body weight also occur seasonally in some animals such as hibernators, psychiatrists have looked to those species with annual cycles for working models of seasonal affective disorder. Work on hamsters is particularly interesting. In these rodents, winter conditions such as regression of gonads and weight gain can be brought on by short photoperiods (lengths of day-

light) and prevented by maintaining long photoperiods. Because light suppresses secretion of pineal melatonin, the duration of elevated levels of this hormone is thought to mediate the seasonal changes in hamsters³. By analogy, psychiatrists have tried to extend the photoperiod in winter to prevent or ameliorate seasonal affective disorder in patients⁴. This treatment was stimulated by the important discovery that levels of melatonin in humans can also be suppressed by sufficiently bright light — the intensity of bright daylight (about 2,000 lux)⁵.

It became evident at the meeting that light therapy administered under psychiatric supervision can be highly successful. In some studies as many as 80 per cent of patients were helped by sitting in front of a large, full spectrum, bright light panel in the morning for 2 hours directly preceding the onset of winter daylight. Success was reported by groups working throughout the United States and in Switzerland.



Patient receiving light therapy at the National Institute of Mental Health, Bethesda, Maryland (courtesy of N.E. Rosenthal).

* Seasonal Affective Disorders, workshop sponsored by the National Institutes of Mental Health, Bethesda, Maryland, 4–5 May 1987.

Improvement in clinical mood ratings typically occurs within a week of initiation of daily phototherapy, and relapse following withdrawal of therapy is equally fast. Ironically, it seems likely that the mechanisms for these therapeutic effects are very different from those operating in the photoperiodic animal models that initially inspired the theorizing about seasonal affective disorder and its treatment with bright light. One problem is that light therapy is reported to be clinically effective even if presented during winter daylight hours at times when it does not extend the photoperiod, and when the circadian rhythm of melatonin is already low⁶. Also, attempts to reverse the therapeutic effects of a lengthened photoperiod by oral administration of melatonin, and to mimic the therapeutic effects of light by extending the duration of daily suppression of melatonin secretion with the β -adrenergic blocker atenolol, have both been largely unsuccessful⁷. Another question related to a photoperiodic melatonin mechanism in phototherapy is whether light previously considered too dim to suppress the hormone⁵ can be clinically helpful⁸, and one must also take into account how infrequently an average urban dweller actually experiences the changes in duration of daylight above 2,000 lux⁹.

An alternative explanation of the therapeutic effect of bright light does not invoke a photoperiodic mechanism: light administered at the appropriate time of day (morning) could correct abnormal, phase-delayed circadian rhythms by its ability to entrain them¹⁰. This explanation also has difficulty explaining reports⁶ that phototherapy administered at midday or in the evening can be clinically effective. However, there was considerable debate at the workshop over how well founded these reports are. A third possibility is that phototherapy may merely induce a direct effect on some neurophysiological process which corrects an underlying abnormality (N.E. Rosenthal, personal communication). Finally, for any explanation, the extent of placebo effects remains another critical unresolved issue.

Most animal researchers were sceptical whether there is an appropriate animal model for seasonal affective disorder. Yet the search for photoperiodic and hibernating animal preparations has stimulated imaginative clinical research by helping formulate hypotheses and providing paradigms for testing. This has resulted in identification and successful treatment of a depressive syndrome which, on the basis of surveys in the New York area, may not be trivial in terms of its prevalence and socioeconomic impact.

The non-pharmacological nature of phototherapy is reason enough for trying to discover the nature of its action and its possible usefulness in treating some non-seasonal affective disorders^{11,12}. Whatever

the mechanisms of seasonal effective disorder and phototherapy eventually turn out to be, basic animal research has made a sizable contribution to the differential diagnosis and treatment of a mood disorder. □

1. *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R)* 3rd edn (May 1987).
2. Rosenthal, N.E. *et al. Archs gen. Psychiat.* **41**, 72–80 (1984).
3. Carter, D.S. & Goldman, B.D. *Endocrinology* **113**, 1261–1267 (1983).
4. Rosenthal, N.E. *et al. Am. J. Psychiat.* **142**, 163–170 (1986).
5. Lewy, A.J. *et al. Science* **210**, 1267–1269 (1980).
6. Wehr, T.A. *et al. Archs gen. Psychiat.* **43**, 870–875 (1986).
7. Rosenthal, N.E. *et al. J. neural Transm.* **21** (Suppl.), 257–267 (1986).

8. Wirz-Justice, A. *et al. Acta Psychiatr. Scand.* **74**, 193–204 (1986).
9. Savides, T.J., Messin, S., Senger, C. & Kripke, D.F. *Physiol. Behav.* **38**, 571–574 (1986).
10. Lewy, A.J., Sack, R.L., Miller, S. & Hoban, T.M. *Science* **235**, 352–354 (1987).
11. Kripke, D.F. *Ann. N.Y. Acad. Sci.* **453**, 270–281 (1985).
12. Lewy, A.J. *et al. J. neural Transm.* **21** (Suppl.), 311–322 (1986).

J. D. Hallonquist is in the Department of Psychiatry and the Research Institute at Mount Sinai Hospital, Toronto, Ontario, Canada M5G 1X5, and the Departments of Psychiatry and Zoology at the University of Toronto; N. Mrosovsky is in the Departments of Zoology, Psychology and Physiology at the University of Toronto, Ontario, Canada M5S 1A1.

History of optics

John Elliot MD (1747–1787)

J.D. Mollon

The unfortunate Mr. Elliot is a character on whom every eye should be directed with charity: — he is one of the first geniuses this or any other country ever produced.

SUCH was the judgement of *The Daily Universal Register* (later to become *The Times*)¹; and for a few brief weeks in the summer of 1787, the name of John Elliot, MD, was known to every newspaper reader. His trial is sometimes recalled in anthologies of crime², but by historians of science he has been largely neglected³ or else confounded with his contemporary, Sir John Eliot, Bart (1736–1786)⁴. Two hundred years after his death it is interesting to examine how his experiments in physiological optics led him to fundamental insights in physical optics.

At the age of fourteen, we are told^{5–7}, Elliot was bound apprentice to an apothecary in Spitalfields, and at the expiry of his time he joined a large practice in Cheapside. It was during this period that he established a romantic attachment to Miss Mary Boydel, niece of the celebrated Alderman Boydel.

At first, if we are to believe Elliot's own account⁷, Miss Boydel encouraged him but later she rejected him, whereupon he left London. By 1780, however, Elliot had returned and set up in business in Carnaby market. In that year he published his *Philosophical Observations on the Senses*⁸, which anticipates Johannes Müller's *Doctrine of Specific Nerve Energies*⁹ and was known to Müller in translation¹⁰.

To understand Elliot's insights into sensory physiology, recall that most eighteenth-century writers supposed that vibrations of the aether or of the air were directly communicated to

the optic and auditory nerves and were faithfully transmitted to the sensorium, the place where sensations were aroused (see refs 11, 12). What was lacking was the concept of a transducer, a sensory receptor tuned to only a limited part of the range of physical frequencies. It can be argued that this inadequate physiological model held back the understanding of physical optics in at least two ways: first, few suspected the existence of infrared and ultraviolet radiation, radiation for which we have no transducers; and second, many physicists, knowing that most colours could be produced by mixing three, believed that there were only three kinds of light¹³.

In a series of masochistic experiments, Elliot showed that mechanical stimulation of the eye or the ear could produce a variety of specific sensations, the sensation being always appropriate to the modality stimulated. He rightly concluded that there must be transducers in our sense organs, different ones for the different frequencies — or to use his words⁸, that

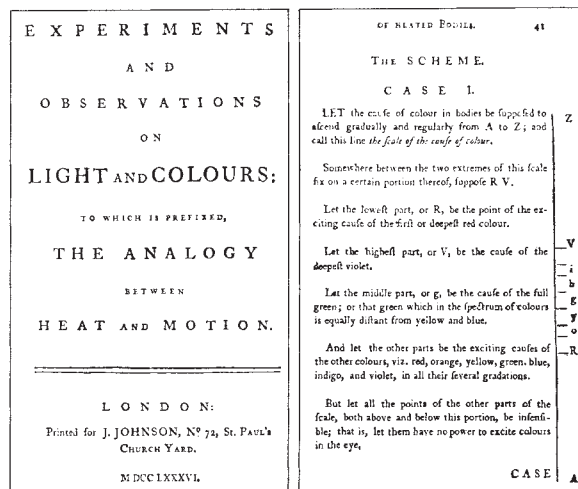


Fig. 1 Title page (left) and text (right) from a monograph of 1786 in which Elliot makes explicit the possibility of optical radiation beyond the limits of the visible spectrum. (Reproduced by permission of Edinburgh University Library).



Fig. 2 A sketch (by Ozias Humphry, RA) entitled *Youth* and thought to be a portrait of Miss Mary Boydell.

... there are in the retina different times of vibration liable to be excited, answerable to the time of vibration of different sorts of rays. That any one sort of rays, falling on the eye, excite those vibrations, and those only which are in unison with them ... And that in a mixture of several sorts of rays, falling on the eye, each sort excites only its unison vibrations, whence the proper compound colour results from a mixture of the whole.

Two years later¹⁴ he elaborates:

We are therefore perhaps to consider each of these vibrations, or colours in the retina, as connected with a fibril of the optic nerve. That the vibration being excited, the pulses thereof are communicated to the nervous fibril, and by that conveyed to the sensory, or mind, where it occasions, by its action, the respective colour to be perceived.

Elliot nowhere says that there are only three retinal resonators and so he never took the small, final step to understanding why all colours can be produced by trichromatic mixture. Conversely, his contemporary, George Palmer, had the idea of three receptors^{15,16}, but did not grasp that the physical variable underlying hue was a continuous one. It remained for the two hypotheses — physiological trichromacy and continuous variation in physical frequency — to be put together by Thomas Young, who was certainly familiar with Elliot's ideas¹⁷.

But Elliot's physiological model did lead him to a physical insight equal to Young's insight into trichromacy. The relationship between light and heat had been discussed by several eighteenth-century authors, but James Hutton is often held to be the first clearly to suggest (in 1794) that there might exist radiation of low refrangibility that had the power of heating but had little power of exciting the

organ of sight¹⁸. William Herschel's empirical study¹⁹ of the infrared, published in 1800, is well known.

In fact, the concepts of the infrared and ultraviolet were introduced explicitly in 1786 in an anonymous work entitled *Experiments and Observations on Light and Colours* (Fig. 1). From internal evidence, and from a manuscript rejected, but preserved, by the Royal Society⁷ we can be sure that the author was John Elliot. Down the side of one page, Elliot draws what may be the first representation of a spectrum extended in both directions beyond the visible region (Fig. 1). His text develops an elaborate analogy between heat and motion, and introduces something very close to the concept of colour temperature:

If the cause of colour be supposed to be simple (for example, if it is supposed to consist in vibration, which may be increased in swiftness, or quickness of return) and to ascend uniformly from A [see Fig. 1], it will be invisible till it arrives at the lowest part of the assumed portion, or R ... It will then begin to become sensible under the form of the deepest red colour. This colour will continually vary, passing gradually ... through all the degrees of red, orange, yellow, green, blue, indigo and violet ... In the light emitted by shining bodies, some colours abound more than others; but the inequality is regular; and the predominant colour varies with the heat according to the foregoing law.

Elliot himself makes explicit the relationship between his earlier physiological model and his physical insights:

... He [a writer on this subject] therefore suggests that the rays of light excite colours in us only by the mediation of these internal colours. From whence it would follow, that if there are rays of light which have no answerable colours in the eye, those rays cannot be visible ...

Elliot grasped, as Herschel later did not²⁰, that the difference between visible and invisible radiation lies in the spectral sensitivity of our eye and that the two forms of radiation are not qualitatively distinct.

Elliot also deserves a place in the early history of spectroscopy. In simple experiments, he observed heated bodies by means of a small aperture and prism:

As the body in the third experiment cooled, it was pleasant to observe how, by degrees, the violet first, and then the indigo, blue, and the other inferior colours vanished in succession, as if the spectrum were contracting itself towards its inferior part; and how the centre of the range seemed gradually to move from orange to red, and at length beneath it, as if it sunk into the insensible part below R in the scheme, the superior part following it, till the whole range was out of sight ...

But Elliot's scientific career was ending. Into his shop one day came Mary Boydell with a lady companion. Thereupon,

by Elliot's account⁶, both he and Miss Boydell swooned. Their relationship was re-established, but Miss Boydell soon rejected Elliot again, and he was once more consumed by bitter melancholy. He bought two brace of pistols. One brace — or so his attorney claimed at the trial — he filled with shot, and the other with blank shot, his intention being to discharge the blank shot at Miss Boydell and then shoot himself dead at her feet²¹. On 6 July 1787, near Leicester Square, Elliot came up behind Miss Boydell, arm in arm with her companion, Mr Nichol. Elliot fired at Miss Boydell, but before he could shoot himself he was seized by Nichol.

At his trial at the Old Bailey, the prosecution insisted that the pistols had been loaded — and that Miss Boydell had been saved only by her whalebone stays. Her scorched dress was produced in court (*The World, Fashionable Advertiser* for 17 July tells us it was "of very elegant white muslin, spangled with gold"). The defence claimed that the pistols had not been loaded; and that in any case Elliot was of unsound mind. Some of his scientific writings were adduced as proof of his insanity. The jury acquitted him, but the Recorder committed him to Newgate nevertheless, to be tried for assault. Elliot entered on a hunger strike and was dead by 22 July. The *Daily Universal Register*, ever on his side, held that he died "of what is commonly stiled a broken heart". □

1. *The Daily Universal Register* London, 13 July, 1787.
2. Cook, C.T. *The Complete Newgate Calendar* 4, 165–168 (Navarre Society, London, 1926).
3. Gillispie, C.C. *Dictionary of Scientific Bibliography* Vol. 4 (Charles Scribner's Sons, New York, 1971).
4. Cantor, G. *Optics After Newton* (Manchester University Press, 1983).
5. *The London Review* pp 44–45, July 1782.
6. A Narrative of the Life and Death of John Elliot, M.D. containing an account of the rise, progress, and catastrophe of his unhappy passion for Miss Mary Boydell; a review of his writings. Together with an Apology, written by himself (Ridgway, London, 1787).
7. Partington, J.R. & McKie, D. *Ann. Sci.* 6, 262–267 (1941).
8. Elliot, J. *Philosophical Observations on the Senses of Vision and Hearing* (Murray, London, 1780).
9. Müller, J. *Handbuch der Physiologie des Menschen* Buch 3 (Hölscher, Coblenz, 1834).
10. Elliot, J. *Ueber die Sinne* (Leipzig, 1785).
11. Newton, I. *Opticks* 4th edn, Queries 12, 13 (Innys, London, 1730).
12. Dagoty, J., Père *Exposition Anatomique des Organes des Senses, Jointe à la Neurologie Entière du Corps Humain* (Demonville, Paris, 1775).
13. Mollon, J.D. in *Images and Understanding* (eds Weston-Smith, M., Barlow, H.B. & Blakemore, C.B.) (in the press).
14. Elliot, J. *Elements of the Branches of Natural Philosophy connected with Medicine* (Johnstone, London, 1782).
15. Walls, G. *J. hist. Med.* 11, 66–96 (1956).
16. Mollon, J.D. *Actes du Seme Congrès de L'Association Internationale de la Couleur* Vol. 1 (Association Internationale de la Couleur, Paris, 1985).
17. Young, T. *A Course of Lectures on Natural Philosophy and the Mechanical Arts* (London, 1807).
18. Hutton, J. *A Dissertation upon the Philosophy of Light, Heat and Fire* (Cadell, Edinburgh, 1794).
19. Herschel, W. *Phil. Trans. R. Soc. Part I* 255–283; 293–326; 437–538 (1800).
20. Lovell, D.J. *Isis* 59, 46–60 (1968).
21. Hodgson, E. *The Trial of Doctor John Elliot; for Feloniously Shooting at Miss Mary Boydell*, (Old Bailey Sessions paper. No. 6, Part 4) (Hodgson, London, 1787).

J.D. Mollon is in the Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK.

Phosphotransferase sequence homology

SIR—I wish to report a remarkable sequence homology which is found in a number of different enzymes all of which confer antibiotic resistance by acting as phosphotransferases. A sequence with many similar features is also found in a number of different protein kinases.

The homologies were found in the course of developing a programming system for the analysis of sequence patterns. Details will be reported elsewhere, but, briefly, programs are written in a specially-designed string-processing language, for which there is a C interpreter. Databases can be searched for matches to templates composed of elements of different types: thus, X is matched only by X; <XYZ>, by X or Y or Z; <> by anything; and <.XY>, by anything excepting X and Y. Gaps are not allowed. A tolerance on the match can be specified to find not only the perfectly matched sequence but also sequences with one or more errors.

The sequences are shown in the accompanying figure. There are two regions separated by a variable number of amino acids; the first region was used to generate a search template. Originally this was taken from the aminoglycoside phosphotransferases (see figure legend for details) aph.1 to 5 and the template used was H<>D:h<><> <>N:h:h:h, where :h stands for <FLIMVWY> (single letter code), the large hydrophobic amino acids. This template found the viomycin phosphotransferase. The homology between the whole region and that of the aph enzymes is very striking, and particularly so because there is no other sequences resemblance between these enzymes. A similar sequence is found in the hygromycin phosphotransferases, which share no other homologies either with each other or with the other phosphotransferases. The streptomycin phosphotransferases show some deviation from the

template, in that they do not have 3 hydrophobic amino acids following the asparagine. The template finds all of the protein kinases. With the exception of aph.6, all the sequences have the invariant asparagine, and all have the invariant pair of aspartic acids.

Most of the antibiotics are aminoglycosides, but viomycin is not, and the protein kinases have serine, threonine or tyrosine in peptide chains as their substrates. The conservation of the sequence suggests that this is a part of the protein that is involved in a common function; the obvious possibility is the binding of ATP and the phosphorylation of the substrate. The common pair of aspartic-acid residues could be involved in the binding through Mg^{2+} of the phosphate groups of ATP, and even in their activation. These and other possibilities are open to test by genetic manipulation of the sequences.

Other common sequence motifs have been reported. There may be many more and I am now engaged in trying to extract these automatically and exhaustively to see what sort of order this imposes on the database.

SYDNEY BRENNER

MRC Unit of Molecular Genetics,
MRC Centre, Hills Road,
Cambridge, CB2 2QH, UK

- Gray, G. & Fitch, W. *Molec. Biol. Evol.* **1**, 57–66 (1983).
- Herbert, C.J., Giles, I.G. & Akhtar, M. *FEBS. Lett.* **160**, 67–71 (1983).
- Beck, E., Ludwig, G., Auerswald, E.A., Reiss, B. & Schaller, H. *Gene* **19**, 327–336 (1982).
- Oka, A., Sugisaki, H. & Takanami, M. *J. molec. Biol.* **147**, 217–226 (1981).
- Thompson, C.J. & Gray, G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5190–5194 (1983).
- Ferretti, J.J., Gilmore, K.S. & Courvalin P. *J. Bact.* **167**, 631–638 (1986).
- Bibb, M.J., Ward, J.M. & Cohen, S.N. *Molec. gen. Genet.* **199**, 26–36 (1985).
- Zalacain, M. et al. *Nucleic Acids Res.* **14**, 1565–1581 (1986).
- Gritz, L. & Davies, J. *Gene* **25**, 179–188 (1983).
- Kaster, K.R., Burgett, S.G., Rao, R.N. & Ingolia, T.D. *Nucleic Acids Res.* **11**, 6895–6911 (1983).
- Mazodier, P., Cossart, P., Giraud, E. & Gasser, F. *Nucleic Acids Res.* **13**, 195–205 (1985).
- Tohyama, H., Okami, Y. & Umezawa, H. *Nucleic Acids Res.* **15**, 1819–1833 (1987).
- Schwartz, D.E., Tizard, R. & Gilbert, W. *Cell* **32**, 853–861 (1983).
- Reimann, E.M. et al. *Biochemistry* **23**, 4185–4192 (1984).
- Van Beveren, C. *Nature* **289**, 258–262 (1981).

```

aph.1 V F S H G D L G D S N I F V G -- D G -- K V S G F I D L G R S G R A D K W Y D I A F C V R S I
aph.2 V F A H G D Y C A P N L I I D -- G E -- K L S G F I D C G R A G V A D R Y K D I S L A I R S L
aph.3 V V T H G D A C L P N I M V D -- N G -- R F S G F I D V G R L G V A D R Y Q D I A L A T R D I
aph.4 V V T H G D F S L D N L I F D -- E G -- K L I G C I D V G R V G I A D R Y Q D L A I L W N C L
aph.5 V V C H G D L C P N N V L L D -- P G T C R V T G V I D W G R L G V A D R H A D I A L A S R E L
aph.6 C L C H N D F S C N H L L L D -- G N N -- R L T G I I D F G D S G I I D E Y C D F I Y L L E D M

vph.1 V V -- H G D L G G E N V L W E T V D G V P R M S G V V D W D E V G I G D P A F D L A A I G A S Y

hph.1 L V -- H A D F G S N N V L T D L A A T -- R I T A V I D W S E A M F G D S Q Y E F A N I F F W R
hph.1 F V -- H G D L H G T N I F V D -- N G -- E V T G I V D F T D V Y A G D S R Y S L V Q L H L N A

sph.1 P L -- H G D L H H E N V D L F -- G D -- R G W L A I D P H G L L -- G E R T F D Y A N I F T N P
sph.2 V L -- H W D L H Y E N V D A A -- E A -- E P W L A I D P E P L V -- A D P G F D Y L W F A L D T G

prk.1 Y V -- H R D L R A A N I L V G -- E N -- L V C K V A D F G L A R L I -- -- E D N E Y * T A R Q G
prk.2 I V -- H R D L K P E N I L L D -- D D -- M N I K L T D F G F S C Q L D P G E K L R E V C G T P
prk.3 I L -- H L D L K P A N I L I S -- E Q -- D V C K I S D F G C S Q K L Q D L R G R Q A S P F N I

```

Enzymes: aph. (3'-aminoglycoside phosphotransferases): 1, *Streptococcus faecalis*¹; 2, *Bacillus circulans*²; 3, Tn5³; 4, Tn903⁴; 5, *Streptomyces fradiae*⁵; 6, 2'-aminoglycoside phosphotransferase⁶; vph. (vio-mycin phosphotransferase): 1, *Streptomyces vinaceus*⁷; hph. (hygromycin B phosphotransferase): 1, pJR225 (*E. coli*)⁸; 2, *Streptomyces hygroscopicus*^{9,10}; sph. (streptomycin phosphotransferase): 1, Tn5¹¹; 2, *Streptomyces griseus*¹²; prk. (protein kinases): 1, src (rous sarcoma virus)¹³; 2, gamma subunit of phosphorylase kinase (bovine)¹⁴; 3, v-mos mouse sarcoma virus¹⁵.

Electric charge of the neutrinos from SN1987A

SIR—The detection of nearly a dozen neutrinos, of 7 to 35 MeV in energy, coming presumably from the 1987 supernova in the Large Magellanic Cloud (LMC), 50 kpc away, all within some ten seconds¹, has been interpreted by several people as a proof that electron anti-neutrinos have a mass no larger than about 10eV.

Here we want to point out that, even if their mass is negligible, the bunching in time of the neutrinos indicates that their charge, q , is smaller than about 10^{-17} times the charge of the electron. If q were larger than this limit, the galactic magnetic field would lengthen their paths, and neutrinos of different energy could not arrive on the Earth within a few seconds of each other, even if emitted simultaneously by the supernova.

If B (tesla) is the uniform magnetic field normal to the path of the neutrino of charge q (in units of electronic charge e) and energy E (GeV), the radius of curvature of the trajectory (in metres) is

$$r = \frac{E}{0.3Bq}$$

The arc, $r\theta$, followed by the neutrino is longer than the straight path, s , by

$$\delta s = r\theta - 2r \sin \frac{\theta}{2} \approx s \frac{\theta^2}{24} \quad (\theta = \frac{s}{r} \ll 1)$$

so

$$\frac{\delta s}{s} \approx \frac{1}{24} \frac{(0.3Bs q)^2}{E^2}$$

The difference in length of the trajectories of two neutrinos of average energy E , differing in energy by ΔE , is then

$$\frac{\Delta s}{s} = \frac{\Delta t}{t} = \frac{1}{12} \frac{(0.3Bs q)^2}{E^2} \frac{\Delta E}{E}$$

where $\Delta t = s/c$ and Δt is the difference in the arrival time of the neutrinos on the Earth.

In conclusion,

$$\frac{q}{e} \leq \frac{\sqrt{12}}{0.3} \frac{E}{Bs} \left(\frac{\Delta t/t}{\Delta E/E} \right)^{1/2} = 3.8 \times 10^{-12} \frac{(E/10 \text{ MeV})}{(s/10 \text{ kpc}) (B/1 \mu \text{G})} \left(\frac{\Delta t/t}{\Delta E/E} \right)^{1/2}$$

For the neutrinos from the LMC supernova, $E \approx 15$ MeV, $\Delta E/E \approx 1/2$ and $\Delta t \leq 5$ s.

Consider two cases: (1) an intergalactic field $B = 10^{-3} \mu\text{G}$ over the whole path, $s = 50$ kpc, $t = 5 \times 10^{12}$ s; (2) the galactic field, $B = 1 \mu\text{G}$, over a distance $s = 10$ kpc, $t = 1 \times 10^{12}$ s. In the first case, we find $q/e \leq 2 \times 10^{-15}$, and in the second, $q/e \leq 2 \times 10^{-17}$. (These examples consider homogeneous fields. A number, n , of randomly oriented inhomogeneities would increase our estimates by a factor of $\sqrt{n}$. In the

case of the Galaxy, however, the fields of 2–3 μG are not random, but run along the galactic arms.)

Both limits are substantially smaller than the limit, $q/e \geq 10^{-13}$, obtained by Bernstein and Ruderman² from arguments on solar energy losses. It is indeed remarkable that from these first ever observations of neutrinos from a supernova such stringent limits on their charge can be obtained.

GUIDO BARBIELLINI
GIUSEPPE COCCONI

CERN,
CH 1211 Geneva 25,
Switzerland

1. Hirata, K. *et al.* *Phys. Rev. Lett.* **58**, 1490 (1987).
2. Bernstein, J. & Ruderman, M. *Phys. Rev.* **132**, 1227 (1963).

Growth transformation by *v-myc*

SIR—A recent report by Casalbore *et al.*¹ showed that the avian retrovirus MC29, which contains only the *v-myc* oncogene, stimulates the growth of neuroretina (NR) cells from 7-day-old embryos, under conditions where uninfected cells can be cultivated for 10–15 population doublings. The authors conclude that the *v-myc* oncogene is sufficient to induce “growth transformation” of chick NR cells and that “their results differ considerably from the experiments of Bechade *et al.* (ref. 2)”.

Indeed, we have reported that viruses containing only the *v-myc* oncogene fail to induce proliferation of normal NR cells from 7-day-old chick embryos which, in our experimental conditions, do not multiply *in vitro*, while oncogenes such as *v-src*³ and *v-mil*⁴ induce rapid and massive proliferation of these cells. In the experiments of Casalbore *et al.*, the *v-myc* gene appears to enhance the growth rate of NR cells that are already undergoing cell division. Therefore, we believe that their results support our conclusion that *v-myc* is not sufficient to induce NR cell multiplication, because it depends on specific culture conditions, that are not required by *v-src* or *v-mil*, in order to alter the growth properties of these cells.

Yet, even under these “optimized culture conditions in which efficient virus spread allows a rapid uniform viral infection” the authors showed that only 30% of multiplying NR cells release MC29, while over 80% of NR cells infected with MH₂, containing both the *v-mil* and *v-myc* oncogenes, were found to produce virus. To explain these differences, they suggest that MC29-infected NR cells are more sensitive to environmental effects that would delay expression of the *v-myc* protein. A simpler and more obvious explanation would be that the concomitant expression of *v-mil*, in MH₂-infected NR cells, allows synthesis of the *v-myc* gene product in the majority of NR cells.

Casalbore *et al.* conclude that their results “suggest a novel experimental approach to generating clonal strains from short-lived early neuronal precursors otherwise not easily amenable to adequate characterization”. Indeed, we have already reported that it is possible to generate clonal strains that exhibit typical neuronal properties with quail NR cultures transformed by Rous sarcoma virus^{4,5}.

GEORGES CALOTHY

Institut Curie-Biologie,
91405, Orsay Cedex, France

BERNARD PESSAC

Centre de Biologie Cellulaire, CNRS,
94205 Ivry, France

1. Casalbore P. *et al.* *Nature* **326**, 188–190 (1987).
2. Bechade C. *et al.* *Nature* **316**, 559–562 (1985).
3. Pessac B. & Calothy G. *Science* **185**, 709–710 (1974).
4. Pessac B. *et al.* *Nature* **302**, 616–618, (1983).
5. Crisanti P. *et al.* *EMBO J.* **4**, 1467–1470 (1985).

The primate trade and the origin of AIDS viruses

SIR—There is now¹ compelling evidence for an African origin of the viruses causing AIDS. It seems that the human AIDS viruses are descended from a monkey virus, probably via the passage to man of the simian T-lymphotropic virus (STLV-3), now called SIV, isolated² from healthy African green monkey (AGM) *Cercopithecus aethiops*, which causes an AIDS-like disease in captive rhesus macaque monkeys³. Thus, an apparently harmless monkey virus has probably given rise to a human virus which has evolved into the HIVs, but it remains difficult to explain why human HIV seropositivity is recognized in Africa only after the 1950s.

Is it possible that these events are linked with the beginning, in the 1950s, of a massive trade of monkeys from Africa to western countries, mainly the United States, coinciding with the beginning of tissue-culture technology? In the 1950s, the introduction of tissue cultures in research into human enteroviruses, and in studies of the preparation and control of polio vaccines, caused a massive request for monkeys, and many primate stables were created where different species of monkeys often lived together. AGM has been one of the monkeys used most for kidney cultures for enterovirus studies. This caused an unprecedented human manipulation of AGM by Africans involved in the capture and maintenance of these monkeys, and stables and laboratory personnel of western countries. All this might have vastly increased the odds of an accidental passage of SIV from AGM to other monkeys and to humans. The lag in the appearance of widespread human infection might have been due to the adaptation of the virus to its new host and to the behavioural characteristics of people in the different countries.

Finally, the susceptibility of rhesus

monkeys and other primates to SIV raises the concern of the safety conditions used for the selection and maintenance of monkeys entering primate stables in general, and of those monkeys used by industrial polio vaccine suppliers in particular; taking into account that blood-cell contaminants (adherent mononuclear cell clusters) are always present in primary *in vitro* monkey kidney cell cultures.

SERGIO GIUNTA
GIUSEPPE GROPPA

Unità di Ricerca Corrente in Genetica
e Sezione di Virologia del Laboratorio
Clinico INRCA,
Ancona, Italy

1. Newmark, P. *Nature* **324**, 611 (1986).
2. Kanki, P.J., Alroy, J. & Essex M., *Science* **230**, 951–954 (1985).
3. Daniel, M.D. *et al.* *Science* **228**, 1199–1204 (1985).

Contingent contagious constraints

SIR—Hoyle and Wickramasinghe¹ present arguments in the cometary debate which are, “contingent on constraints that they themselves have imposed”. Where they have argued that, “the available epidemiological evidence is consistent with the model where the causative agent (or a trigger for it) is airborne, “they have analysed epidemics in an eccentric manner of their own devising which is entirely innocent of the techniques of mathematical epidemiology² — an approach which would be strange enough were the authors non-numerate but which is truly extraordinary for scientists with such considerable mathematical powers³”.

A key feature of contagious processes is that they can exhibit great variability. This is especially likely to be the case when the values of parameters are near threshold, and processes in which variances are greater than means are not uncommon. In their analyses², however, Hoyle and Wickramasinghe repeatedly rely on the binomial distribution (for which the variance is less than the mean) and their claims to demolish theories of person to person transmission on this basis are no more impressive than would be those of a medical statistician claiming to have overturned classical physics because he had discovered that force was not proportional to velocity.

S.J. SENN

Dundee College of Technology,
Bell Street, Dundee, UK

1. Hoyle, F. & Wickramasinghe, C. *Nature* **327**, 664 (1987).
2. Hoyle, F. & Wickramasinghe, N.C. *Diseases from Space* (Dent, London, 1979).
3. Senn, S.J. *New Scientist*, **92**, 244–246 (1981).

Erratum

Reference 2 in J.H. Hecht's letter in *Nature* **328**, 765: 1987, should have been: Saslaw, W.C. & Gaustad, J.E. *Nature* **221**, 160–162 (1969).

A buyers' market

Walter Gratzer

Communicating Science to the Public. Edited by David Evered and Maeve O'Connor. Wiley: 1987. Pp.214. £28.95, \$49.95.

THERE WAS ONCE an entrancing headline in a British newspaper, which ran: *Incest more common than thought in United States*. Well, the news is that it may be true. At least on the evidence deployed by Alan J. Friedman, who brings these symposium proceedings to their sombre close, unreason reigns, and education, by inexorable inference, has been a failure. Fifty-five per cent of American teenagers believe for instance that astrology works. As for the British, *seventy-five* per cent are persuaded that it is a scientific discipline. This, it appears, is often held to follow (well, who can doubt it?) from the well-known circumstance that astrologers use computers. Nor is it only the common multitude that holds such convictions, for President Reagan owns to being guided by his horoscope. "I believe you'll find", he says, "that 80% of the people in New York's Hall of Fame are Aquarians". Friedman has even discovered a member of the Faculty of Columbia University (admittedly a psychiatrist) with closely similar opinions.

Ours is an age that will chiefly be remembered by history for its scientific and technological achievements. Why then should it matter if a large proportion of our populace believes that oysters engender lust and that the Milky Way is God's daisy chain? The learned group of experts, who were brought to the Ciba Foundation to consider this and similar questions, return no very clear answer. Dr Miller from the Public Opinion Laboratory of Northern Illinois University quotes one I.C. Davis, writing 50 years ago, that he who has acquired the scientific attitude will:

(1) show a willingness to change his opinion on the basis of new evidence [though this was C.M. Bowra's reason for regarding scientists as unpredictable and therefore dangerous on academic committees]; (2) will search for the whole truth without prejudice; (3) will have a concept of cause and effect relationships; (4) will make a habit of basing judgement on fact; and (5) [most stirring of all] will have the ability to distinguish between fact and theory.

W.M. Laetsch, from Berkeley, in a magisterial opening chapter, makes short work of such pompous bunkum, which still often emanates from committees of scientists. One of my own favourite examples comes from Karl Pearson, the geneticist (a calling that seems to predispose to fatuous utterances on social issues): "Modern science, as training the mind to

an exact and impartial analysis of facts, is an education specially fitted to promote citizenship". Not so very long ago it was regarded as self-evident, at least in Britain, that this was uniquely the attribute of a classical education. (And in the United States the introduction of the BSc

"It is in reality a matter of observation that scientists given to public vociferation on subjects not their own will too often cast off the constraints of caution and intellectual honesty that vex them in the laboratory. We all know that no proposition is so foolish or meretricious that at least two Nobel laureates cannot be found to endorse it."

degree drew the following comment from a Dean of Harvard: "It does not guarantee that the holder knows any science, but it does guarantee that he does not know any Latin".) It is in reality a matter of observation that scientists given to public vociferation on subjects not their own will too often cast off the constraints of caution and intellectual honesty that vex them in the laboratory. We all know that no proposition is so foolish or meretricious that at least two Nobel laureates cannot be found to endorse it.

Laetsch's conclusion, which is both wholesome and refreshing, is that ultimately the teaching of science must be defended, like all other forms of education, on the grounds that it is able to give pleasure and enrich life. Now it would, I think, be paltering with the truth to pretend that our activities do as much in the cultural line for the public at large (especially at the price) as say the London Philharmonic playing in the Festival Hall. The standards for a start are lower; most science is too narrow or trivial to give pleasure even to other scientists. It was not always so: in Victorian times scientific discoveries formed the substance of improving after-dinner entertainments for the family; collections of microscope slides and *The Origin of Species* probably reposed on the shelf in most homes with any pretension to culture. More recently, the three-volume popular treatise on biology, by H.G. and G.P. Wells and Julian Huxley, enjoyed enormous success. H.G. took the enterprise very seriously. "The job", he wrote to Julian Huxley, "is an important job; your own researches

and your professional career are *less important*".

What then has gone wrong? Geoff Deehan of the BBC writes entertainingly of the broken-backed and half-hearted attempts by scientists (especially the British) to explain their work to the public. Rutherford said that no theory was to be taken seriously that could not be explained to a barmaid, and this is surely an enduring truth. A recurring theme of the book is that scientists have little interest in repaying society in the same coin. According to Dorothy Nelkin, they see the mass media only as engines of political persuasion, especially when times are hard. Here is a newspaper editor, quoted in her discourse on science and the media:

When NSF money was available easily, you couldn't get a story out of a molecular biologist. Today I get copies of grant applications in the mail with this thing, 'single cure for blank or whatever it might be', circled in red, saying 'we need all the help we can get, fellers'.

Nelkin has another good illustration: a *New York Times* science writer surveyed, evidently with circumspection and objectivity, the state of progress on interferons, and cautioned his readers not to expect miracles. This drew a complaint to the editor from a group of scientists that such a cool evaluation could jeopardize their funding.

The generally unflattering picture of scientists that emerges from these pages is compounded in Alan Friedman's contribution by our image in fact and fiction. Here is a students' eye view:

He neglects his family — pays no attention to his wife, never plays with his children — He has no social life, no other intellectual interests... He bores his wife, his children and their friends... He is always running off to his laboratory. He may force his children to become scientists also.

Oh, well — better perhaps than W.H. Auden:

To the man in the street, who I'm sorry to say
Is a keen observer of life,
The word intellectual suggests straight away
A man who's untrue to his wife.

It is curious perhaps to find a symposium on the communication of science that has no contributions from the great popular communicators, or indeed (except in the discussions) from any scientists. There are, at least to my untutored taste, too many of the kind of sociological effusions that flutter out of Departments of Education like bats out of a barn. But there are some excellent chapters, in between the *longueurs*. It is a worthwhile enterprise and merits a $\beta(+?)$ for earnest good intentions. □

Walter Gratzer is in the Medical Research Council Cell Biophysics Unit, King's College London (KQC), 26-29 Drury Lane, London WC2B 5RL, UK.

Genetic assortment

Jeffrey Williams

Molecular Genetics of Mammalian Cells: A Primer in Developmental Biology.

Edited by George M. Malacinski. Macmillan New York/Collier Macmillan London:1987. Pp.389. \$52, £45.

PITY the plight of an editor of a volume such as this, which attempts to review a massive field of research where new insights and discoveries are accumulating at a dizzying rate. Given a finite limit to the length of the book, overall coverage will be patchy. Add in the normal delay between the time of writing and of publication, in a hardback, glossy format, and perspectives in certain areas will inevitably be outdated. In this particular collection of essays, the editors have compounded the former problem by including several chapters which deal solely with the work of a single group in what can be best described as bywaters of mammalian cell research. Thus the chapters on cell mutants in RNA polymerase II and on drug resistant dihydrofolate reductase genes are worthy but hardly merit the space devoted to them. The latter article could easily have been condensed to a single paragraph in the excellent review of dihydrofolate reductase gene amplification which follows it.

What makes this misuse of space particularly reprehensible, in a self-professed "Primer in Developmental Biology", is the almost complete absence of any discussion of *developmental* regulation of gene expression. Thus important differentiation-inducible systems, such as mouse erythroleukaemia or teratocarcinoma cell lines, are not described. In addition to inconsistency in the choice of subject, the individual articles vary enormously in depth of coverage and intellectual rigour. The chapter on housekeeping genes, and that on DNA methylation, are poorly written, uncritical and add nothing to previous reviews. In contrast two articles on hormone inducible genes, and a review of enhancer elements by the late George Khoury, are very worthy additions to the literature. Even this latter article is, however, but a starting point for further reading because it pre-dates the explosion of information concerning enhancer-binding proteins.

Given the rate of progress in this field, what is the role of such a book? If it is to be purely educational, then it probably will not be of great value. The beginner can read the new generation of text books, which are almost as comprehensive and up to date as some of these articles, and those with some background knowledge can consult the last few years' review pages from leading biological journals. The book

does, however, have other stated aims, and in these it is successful. It adequately conveys, for the non-initiated, a sense of the mechanics and the power of the techniques involved and it highlights several potential applications of mammalian cell genetics, such as the production of commercially important proteins and the correction of human genetic diseases. Add to this the several worthwhile articles

already mentioned and the book does merit inclusion in a departmental library with money left to spend at the end of the financial year. It would not, however, be one I would consider adding to my personal collection. □

Jeffrey Williams is at the Clare Hall Laboratories, Imperial Cancer Research Fund, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3LD, UK.

Bridge of size

Henry Petroski

The Gate: The True Story of the Design and Construction of the Golden Gate Bridge. By John van der Zee. Simon & Schuster: 1987. Pp.380. £14.95, \$19.95.

THE prehistory of a big engineering project can span decades and even centuries, but it is not because our dreams to build and manufacture must wait for our science to catch up. Innovative engineering, especially in the realization of dream bridges and dream machines, has often actually led science and revealed new phenomena of nature to be explained. The words and formulae of now-sophisticated sciences such as structural mechanics and thermodynamics came long after the often ineffable appearance of a grand new bridge carrying a grand new steam locomotive across a grand old expanse of water and air.

The greatest feats of engineering do not arise full blown out of equations; they arise out of imagination, and determination and perseverance. A bridge builder must first overcome not the intricacies of scientific analysis but the machinations of

ferry boat operators who would be left high and dry. He must answer the critics who say there is no firm riverbed on which to found his piers. He must answer the environmentalists who say his bridge will spoil the landscape. He must answer the Navy which says his bridge will impede their ships. And he must answer the engineers who say his design is technically not feasible. A bridge, then, is not a unique solution to a differential equation but a cultural and technological compromise.

The history of the origins, design, construction and continuing symbolic presence of the Brooklyn Bridge is archetypal, and it has been related by David McCullough in his book *The Great Bridge* (1972). Now John van der Zee has told the similarly compelling story of the Golden Gate Bridge on the occasion of the fiftieth anniversary of its opening. Neither author is an engineer, but both have mastered the technical details of their subjects. And both have used their talents as writers to narrate not only the technical but also the human drama involved in bringing the concept of a great bridge to fruition.

Engineering projects necessarily involve a large cast of characters, and van der Zee has portrayed his as deftly as a novelist might. The engineers in this book

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

California dreaming — the Golden Gate Bridge under construction.

come alive as people, with all the faults and foibles associated with the human species. The story of the Golden Gate Bridge is principally the story of its chief engineer, Joseph Strauss, and he is both hero and villain of the piece. Strauss was head of an enormously successful Chicago firm that designed and patented bascule bridges of generally modest span and, at best, unremarkable appearance. One such bridge was built in San Francisco around 1916, and it enabled Strauss and the ambitious and influential city engineer, Michael O'Shaughnessy, to cross paths.

O'Shaughnessy had long thought of a bridge across the Golden Gate, and often talked of the idea with engineers whom he met in the course of city business. Before he met Strauss, the estimates O'Shaughnessy received were mostly of the order of a quarter of a billion dollars and full of caveats. Strauss claimed he could build a bridge for under \$25 million, and in 1921 produced an ungainly design that was priced at \$17 million. The next lowest estimate was still four or five times as high. Soon Strauss had immersed himself in West Coast politics and cleared the way for financing his project.

How Strauss's ugly duckling evolved into the beautiful Golden Gate Bridge is a fascinating tale. It is complete with revelations about how Charles Ellis, a classics scholar and self-taught bridge engineer, really translated Strauss's conceptual design into an engineering reality. The falling out between Strauss and Ellis, resulting in the latter being denied any official credit for his work on the bridge, was true tragedy.

But the history of the bridge itself, which was designed not so long before the ill-fated Tacoma Narrows Bridge, is a case study of personal and technological adventure bordering on hubris. The relative slenderness of the main suspended span of the Golden Gate was surpassed only by that of the Tacoma Narrows, and in high wind the flexibility of the Golden Gate manifested itself in oscillations not unlike those that were observed in the Tacoma Narrows before it collapsed in 1940. The flexibility of the Golden Gate was also made clear on its recent anniversary by its visible deflection under the weight of a quarter of a million confident and celebrating pedestrians. But flexibility is not failure, of course, and the Golden Gate stands today as a triumph of economical engineering. John van der Zee has captured all of this in a fascinating book that shows that the best of cutting-edge engineering is much, much more than science and technology. □

Henry Petroski is a Professor in the Department of Civil Engineering, Duke University, Durham, North Carolina 27706, USA, and author of *To Engineer Is Human: The Role of Failure in Successful Design* (St Martin's Press, 1985).

Green variations

Peter D. Moore

Climate and Plant Distribution. By F. I. Woodward. *Cambridge University Press*: 1987. Pp.174. Hbk £22.50, \$39.50; pbk £8.95, \$14.95.

DURING the course of their evolutionary development, botanists have spent an inordinate amount of time and effort on the question of what determines the geographical ranges of plant species and vegetation types. Climate is generally accepted as a major factor in limiting such ranges, but its precise mode of action is often difficult to disentangle from the complex knot of environmental/physiological interactions. Besides the range of physical variables implied by the term 'climate', there are also such matters as scale of study, biological interactions between species and the time dimension to be taken into account if the plant's relationship to climate is to be adequately examined. It is Woodward's aim to deal with this complex subject in the course of a concise book.

Most authors who have tackled vegetation/plant relationships begin with an explanation of the apparently static, frozen moment of the present and then expand into a consideration of vegetation dynamics and climatic change. Woodward chooses the reverse approach and begins his study with the past and the palaeoecological evidence for climatic changes — an interesting but rather dangerous way into the subject. Immediately one is faced with the possibility of circular argument. Fossil records of past vegetation provide evidence of climatic change, and changing climate has resulted in the modification of vegetation. But Woodward carefully circumvents this problem by spending little time on the evidence of fossil floras and vegetation. Instead he concentrates on more direct measures of climatic change, such as the use of oxygen isotope variations in marine sediments and ice cores as an indication of changing ocean and atmospheric temperatures. He is then in a position to correlate these changes with those observed in the terrestrial fossil record from vegetation and thus avoid circularity in his reasoning. A wise choice of illustrations here has limited consideration to studies where such evidence as pollen data or timberline altitude have been coupled with oxygen isotope data.

Having begun in this way with the time dimension, the consideration of scale clearly has two aspects, spatial and temporal. Of these, the temporal component involves the rate and amplitude of climatic change and the response time of the vegetation. Immediately we are plunged into the problems of population biology, life

cycle and, most particularly, generation time of the species in question. Cyclical variation in climate and the incidence of catastrophe are involved in the sifting out of available species in a similar way. Short generation time plants, such as annuals, will be sensitive to short wavelength cycles (of the order of days or weeks), whereas long-lived trees may be sensitive only to cycles of 15 years or more.

The scene is then set to look at global climate, in other words to introduce the spatial dimension, and this is tackled from an energy balance point of view. But even here the time element once more raises its head and climatic change, whether deterministic (such as the Milankovitch cycles) or stochastic (such as changes in the circumpolar vortex), may be expected to play its part in influencing the spatial pattern of vegetation. Translating these climatic variables into vegetation predictions demands one further area of knowledge, namely the ecophysiological responses of the plants themselves, and emphasis is placed upon temperature (influencing life form) and hydrology (influencing leaf area index) as providing a means of predicting vegetation physiognomy on the basis of climatic information. Predictive models, with quite impressive outcomes, are developed on the basis of these two factors. Finally, such additional topics as the latitudinal variation in species diversity, the limitations imposed on taxa by their dispersal mechanisms and the interaction between species in the final struggle for the occupation of a particular location are looked at.

As with other books in the series, this study is not intended to be introductory; it assumes a considerable background knowledge in ecology and physiology. This, coupled with its brevity, means that the author must keep up a hot pace of conceptual development in the text. There is no point at which one can afford to lose concentration; indeed, many sentences need to be read twice in order to absorb their full meaning and implication. But this is in no way a criticism, rather a compliment to the author for his skilful distillation of information.

My main complaint regarding clarity concerns figure captions, which are far too brief to explain the meaning, let alone the significance, of the information the figures contain. Very many of the figures cannot be appreciated without close reference to the text. This is a fault which could easily be remedied in future editions.

Overall, this book must be regarded as an extremely fresh and novel approach to a well worn subject. It is a stimulating, thought provoking and important contribution to the study of climatic controls on plant distribution. □

Peter D. Moore is in the Department of Biology, King's College, University of London, Campden Hill Road, London W8 7AH, UK.

Current affairs

Peter D. Killworth

General Circulation of the Ocean. Edited by Henry D. I. Abarbanel and W. R. Young. Springer-Verlag: 1987. Pp.291. DM 160.

THERE are probably three main classes of books on science. Textbooks, with a readership of students and researchers; simplified accounts of developments in a field, intended for the intelligent layman or the scientist who wants to broaden his knowledge; and books that assume some understanding of the subject by the reader, and aim to convey the beliefs, rationales and many of the methods of active researchers. A book in this third class is a rare phenomenon. Unless it is written as a monograph by one individual, it requires a lot of coordination by its editors: one man's trivia can be another man's important material. To succeed in the marketplace, such a book must "feel" different from those in the other two classes.

General Circulation of the Ocean lies in this third category. Its aim is to pass on the current thinking about the behaviour of the ocean on long time scales, using the contents of five sets of lectures given by the authors in 1983. The selection of material achieves that aim well. It also, remarkably, gives a clear impression of the authors delivering their lectures at a blackboard, with asides like "I'll come clean at the outset and state my conclusion"; thus the "feel" is there. Some readers may be irritated by this style; I enjoyed it. However, those present at the lectures were given the opportunity to ask questions — the book's reader cannot. There will therefore be readers who wonder what the terms in the table on p.22 conveyed in the paper cited, or who seek in vain for '(4.25a)' on p. 173; or who wonder where Young's equation (3.13), which conflicts with Hendershott's (paragraph 1.6.9), came from (Hendershott's is the correct one.) These are symptoms of the many minor lapses in editing, which together with the atrocious English ("plateauing", "dimensionfull") and poor diagram quality — most are apparently photocopies from papers — leave one unconvinced that enough care went into the book's production. Whether this matters depends on the readership. Experts, dipping into the book, can fill in the gaps without difficulty; students would have more trouble.

I would have liked Niiler's admirable distillation of observations related to theory to have been twice the length; however, the organizers insisted that "the number of lectures be moderately finite"! Pedlosky's survey of classical (1970s)

thermocline theory, plus 1980s updates of layered "ventilated" theory, shows one of the many excellent features of the book: the quality of total honesty with the reader. There is a temptation when writing a paper to conceal under a plethora of algebra the less successful elements of one's model. That is not so here: strong and weak points are discussed together.

Veronis's section on inverse modelling (the attempt to find a "best fit" velocity field given observations of temperature and salinity data) is the shortest in the book. However, in terms of the amount of immediately useful information and advice for the researcher, it weighs in as the heaviest, and is, simply, the best description of inverse methodology I have read. Young's lectures on baroclinic theories of wind driven circulation overlap somewhat with Hendershott's more historical review of barotropic circulation and some pruning could usefully have been done here. Both are good treatments, although the 20 pages allotted to eddy flux parameterisations seems over long for the end results. Unlike the other lectures, both sometimes let pages of mathematics pass by without the occasional physical discussion, which would have been useful.

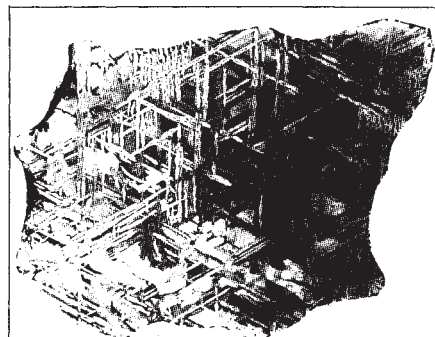
Solid contribution

Colin Bates

Crystalline Semiconducting Materials and Devices. Edited by Paul N. Butcher, Norman H. March and Mario P. Tosi. Plenum: 1986. Pp.645. \$107.40.

ONE of the most popular lines of solid state research in both physics and chemistry is concerned with various types of semiconductors. Apart from their intrinsic scientific interest, semiconductors have many widespread applications which enter into several areas of high technology. New materials are continuously being sought and the literature is full of new results and ideas. Interest originally centred on the bulk materials but more recently the emphasis has moved to the study of ultra thin films and the interface between two materials and superlattices. Another aspect of current research is the careful control of impurities, imperfections and dopants intentionally added during the manufacture.

Such a wide-ranging activity needs some comprehensive base from which adequate but simple models can be built, thus providing the necessary framework upon which the properties of the various materials can be characterized and compared. This book serves admirably the requirement of such a base. It is made up of 16 independently written chapters from 17 contributions from well-known



Cosmic crystals — the Waingaromia meteorite from New Zealand, showing the Widmanstätten pattern formed by intergrowth of kamacite plates within a taenite lattice. The picture is taken from *Meteorites and Their Parent Planets*, by Harry Y. McSweeney Jr, newly published by Cambridge University Press. Price is £15, \$24.95.

But the most important feature of the book is that it fills the reader with a sense of excitement. Ocean circulation studies are exciting and this book lets the reader know it. □

Peter D. Killworth is a Principal Scientific Officer at the Deacon Laboratory, Institute of Oceanographic Sciences, Wormley, Godalming, Surrey GU8 5UB, UK.

European and Japanese universities and laboratories. It begins by discussing the structure of bonds and bands in perfect semiconductors and then goes on to consider thermal and electrical conduction in crystalline structures, followed by a discussion of the effects on the band structure achieved by including shallow and deep impurities centres. After this examination of fundamental work on bulk materials, surface and interface effects are discussed. The last four chapters are concerned with semiconductors as devices. For example, their use in optical communication is discussed in some detail.

In my opinion, this book has succeeded in its attempt to provide a complete account of the basic physics and chemistry of semiconductors. The chapters have been carefully written, although, as is usual in a book of this type, cross-referencing between the chapters is not very frequent. The earlier chapters, being much more fundamental, rely on the more traditional and standard aspects of the basic theory. In contrast, the chapters dealing with the application to devices contain far more references to recent work — up to 1984.

Both graduate research students and more experienced research workers should find the book an invaluable reference source to enable them to understand the more advanced and specialized research paper. □

Colin Bates is a Professor in the Department of Physics, University of Nottingham, Nottingham NG7 2RD, UK.

Chemical processes governing soil and water acidification

J. O. Reuss*, B. J. Cosby† & R. F. Wright‡

* Department of Agronomy, Colorado State University, Fort Collins, Colorado 80523, USA

† Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia 20903, USA

‡ Norwegian Institute for Water Research, Box 333, 0314 Oslo 3, Norway

Strong empirical evidence links deposition of strong acids from the atmosphere to acidification of freshwaters and loss of fish populations. Chemical processes in soils explain the composition of soil solution and surface waters and when coupled with input-output fluxes these processes, in concert with acid deposition, account for observed trends in soil and water acidification. The widespread recent acidification of surface water in eastern North America and Europe is due to the superposition of acid deposition upon natural acidifying processes in soils.

DURING the past several decades, lakes and streams in large regions of northern Europe and eastern North America have undergone acidification and suffered loss of fisheries¹⁻³. These same regions currently receive atmospheric inputs of strong acids at levels several times greater than pre-industrial background levels⁴. There is thus strong empirical evidence linking acid deposition to acidification of surface waters in both time and space^{1,2,5}.

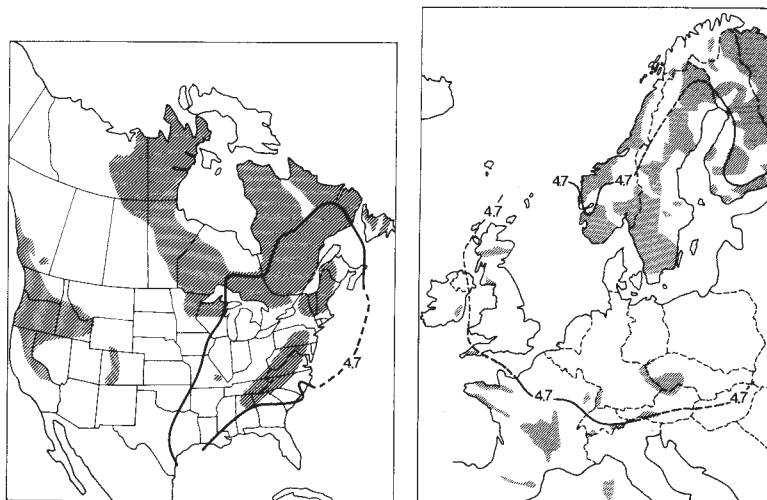
Two factors appear necessary and sufficient to explain the distribution of acidified waters: the inherent sensitivity of the landscape to acidification and acidic atmospheric deposition^{1,2,6} (Fig. 1). Acidification is most pronounced in poorly buffered, low-ionic-strength surface waters. Many thousands of lakes of this type dot the glaciated regions of the Precambrian shields of North America and northern Europe. Such 'sensitive' waters typically drain areas underlain by granitic or other highly siliceous bedrock covered by thin and patchy soils.

Freshwaters in sensitive regions not receiving acidic deposition are generally calcium-magnesium bicarbonate waters with pH levels above 5.5 (ref. 7). Such waters are abundant, for example, in northern Scandinavia^{1,8}, Labrador⁹, northwestern Ontario⁹, and the western United States¹⁰. Sensitive regions receiving acidic deposition, on the other hand, commonly have sulphate-dominated waters that are often acidic and aluminium-rich, with pH levels below 5.5 (ref. 7). Such acid waters are found, for example, in southern Scandinavia¹¹, southern Quebec¹², and the Adirondack Mountains of New York¹³.

In spite of the strong correlation in time and space between the deposition of strong acids and the acidification of freshwaters and resultant loss of fisheries in sensitive regions, the soil-mediated processes by which acidification occurs have only recently been quantitatively described¹⁴. The fact that acidified waters are generally found in areas with acidic soils has led some to question the role of acid deposition and instead argue that changes in the terrestrial ecosystem alone can account for the observed acidification of freshwaters^{15,16}. These sceptics point out that acidic podzolic soils contain acid in amounts equivalent to thousands of years of acid deposition, and thus draw the conclusion that atmospheric acidic deposition is inconsequential compared to naturally produced acid stored in the soils. As we show here, neither the empirical evidence nor basic soil science theory supports this conclusion.

Soil chemical processes lie at the heart of the acidification process; the typical recently acidified lake is a drainage lake which receives only a minor fraction of water directly from the atmosphere. (Seepage and perched lakes, of course, receive significant acid inputs directly from the atmosphere.) Resolution of the apparent paradox by which alkaline waters can be derived from acidic soils in regions not receiving acid deposition, and by which a relatively small amount of strong acid added to an acidic soil can cause a dramatic decrease in the pH of runoff, requires consideration of several key soil-chemical processes. We examine here the influence of several such processes: the adsorption and desorption of anions and cations on soil surfaces,

Fig. 1 Maps of Europe and North America showing areas sensitive to freshwater acidification (from bedrock geology). Sensitive areas are characterized by carbonate-free, highly siliceous bedrock, overburden and soils. Soils in such areas are typically podzolic, acidic and have low base saturation. Freshwaters in these are poorly buffered and of low ionic strength¹. Areas within the pH 4.7 isoline receive precipitation of acidity exceeding the threshold for ecological effects in the most sensitive waters. Acidified freshwaters are largely confined to areas within the pH 4.7 isoline (modified from refs 2, 6).



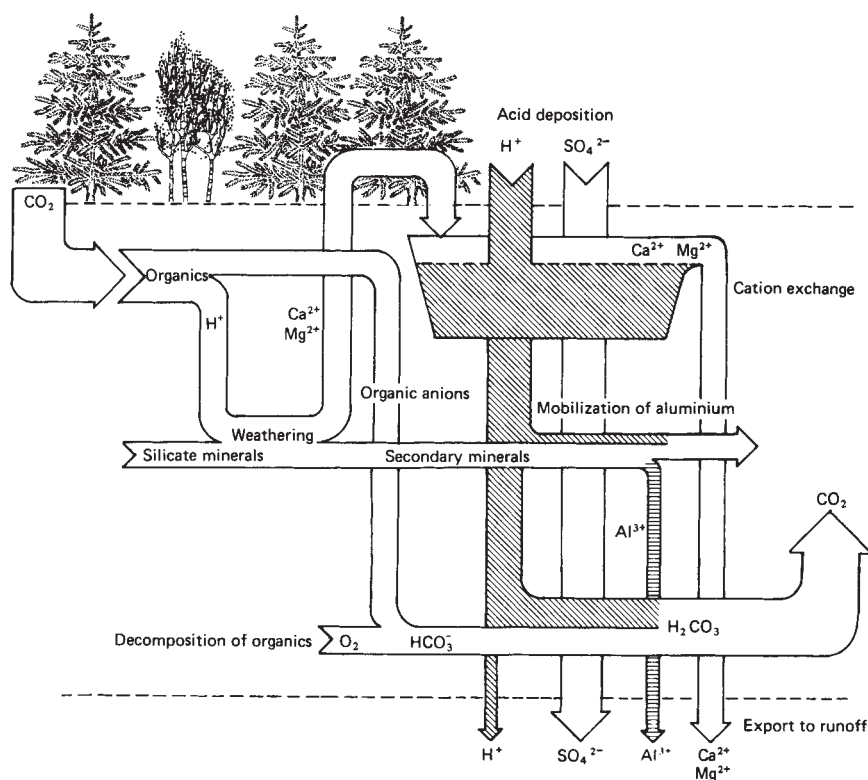


Fig. 2 Schematic view of key processes governing acidification of soil and water. Nitrogen processes are not included; nitrate and ammonium are minor ions in most surface waters. (Modified after ref. 8.)

the dissolution of minerals, and the natural formation of alkalinity in soils (Fig. 2). Although these processes by no means exhaust the possible processes influencing soil and water acidification, they can account for much of the empirical evidence linking acid deposition to water acidification.

Equilibrium relationships

Acid deposition increases the external input of strong acid anions (SO_4^{2-} and NO_3^-), while the accompanying cations are largely H^+ and NH_4^+ . The acid-base relationships of nitrogen processes are complex, but the acidifying effect in soils of the oxidation of NH_4^+ to NO_3^- is well known^{14,17}. The potential of sulphur and nitrogen compounds to acidify surface waters is determined by the flux of strong acid anions SO_4^{2-} and NO_3^- from the soil; the nitrate comes either directly from the atmosphere or from oxidation of atmospheric ammonium. In most areas acid deposition generally contains more sulphate than nitrate. Furthermore most of incoming nitrogen is retained by the terrestrial catchment. Sulphate levels in acidified surface waters are commonly 5–10 times higher than nitrate^{7–9,11,13}, and NH_4^+ concentrations in surface waters are usually minor^{9–11,13}. We thus confine our discussion here to the effect of the SO_4^{2-} ion.

Sulphate concentrations of sensitive surface waters can be accounted for by atmospheric deposition^{9,11,13,18,19}. Increased SO_4^{2-} deposition is usually reflected in elevated concentrations of SO_4^{2-} in soil solution and surface waters. Atmospheric inputs generally far exceed the biological sulphur requirement. Sulphate adsorption may delay the response in sulphate concentrations in runoff to changes in sulphur deposition²⁰. Sulphate adsorption is a concentration-dependent process, commonly described by a Langmuir isotherm²¹. The process may not be simple, nonspecific and completely reversible, however, in which case sulphate concentrations in runoff may exhibit a different response to decreasing deposition than to increasing deposition²⁰.

The capacity to adsorb SO_4^{2-} is apparently related to the processes of soil formation; relatively young forest soils found in northern North America and northern Europe adsorb very little SO_4^{2-} as compared to the more highly weathered soils rich in amorphous sesquioxides common in the southeastern United

States²². Sulphate flux in runoff in these northern areas is generally at steady-state with respect to atmospheric inputs, while SO_4^{2-} flux in runoff in the southeastern United States lies generally well below atmospheric inputs^{14,18,23–25}.

Hydrogen-ion inputs to acid soils will participate in two important processes—cation exchange and release of inorganic Al species through the dissolution of soil minerals. Al^{3+} will tend to displace base cations Ca^{2+} , Mg^{2+} , Na^+ and K^+ from negatively charged exchange sites, because ions of higher valence are preferentially adsorbed²⁶. The net effect of these processes is to provide a buffer for H^+ in soil solution²⁷. As H^+ concentration increases in solution, both processes cause a net removal of H^+ and a net release of base cations and aluminium. The increase in anion concentration in soil solution brought about by acid deposition results in a relatively larger increase in aluminium than base cation concentrations. In acid soils with low base saturation significant amounts of Al^{3+} may remain in solution. The effectiveness of these buffer processes and the

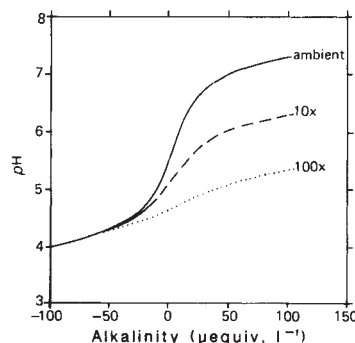


Fig. 3 pH as a function of water alkalinity at ambient ($10^{-3.5}$ atm), 10, and 100 times ambient P_{CO_2} . Elevated P_{CO_2} and pH buffering in soil solutions are ultimately responsible for the pH and alkalinity of surface waters. A pH 5.2 soil solution at ambient, $\times 10$ and $\times 100$ CO_2 levels will have alkalinities of -6, 0, 61 $\mu\text{equiv. l}^{-1}$, respectively. When these waters equilibrate with ambient CO_2 , the pH from the respective soil P_{CO_2} levels will be 5.2, 5.7 and 7.1. Thus, for soil solution at pH 5.2 the resultant surface water pH may vary by 1.7 units, depending on the P_{CO_2} in the soil (from ref. 14).

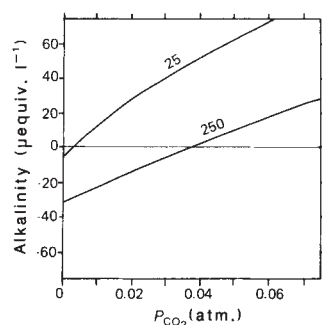


Fig. 4 The effect of strong acid anion concentration (25 and 250 $\mu\text{equiv. l}^{-1}$) and P_{CO_2} on the alkalinity of water in contact with the soil ion exchange complex (from ref. 14). Degassing of soil solution affects the concentrations of bicarbonate and hydrogen ions but not the alkalinity or strong acid anions.

relative amounts of Al^{3+} and base cations brought into solution depends on the aluminium solubility, the ion-exchange selectivity coefficients, and the base saturation of the soil.

The overall importance of the $\text{CO}_2\text{-HCO}_3^-$ equilibrium to both soil and water chemistry is well known, although its role in determining the alkalinity of water draining acid soils has often been overlooked. In acid soils the CO_3^{2-} and OH^- ions can be neglected, and the alkalinity (acid neutralizing capacity) can be defined as

$$\text{alk} = (\text{HCO}_3^-) - (\text{H}^+) - \sum \text{Al}^* \quad (1)$$

where the aluminium term includes Al^{3+} as well as other positively charged inorganic aluminium species. Organic anions of high pK may also contribute to alkalinity; the (H^+) term in our definition includes the acidic effect of organic acids not protonated at pH values above about 5.0–5.5. For surface waters alkalinity is commonly defined as the excess of positive charges over the anions of strong acids²⁸, but by this definition alkalinity is not conserved for water draining acid soils. In soils most H^+ reacts with minerals so that acidity is stored as Al species. In drainage water the reverse reaction acts to buffer increases in pH .

For solutions in contact with soils alkalinity as defined by (1) changes with P_{CO_2} . The product $(\text{H}^+)(\text{HCO}_3^-)$ increases directly with P_{CO_2} , but (H^+) is held relatively constant by the dissolution of soil minerals and the exchange of H^+ and Al species for base cations. Thus when P_{CO_2} increases, (HCO_3^-) increases. This gives rise to increased alkalinity to the extent that this increase in anion concentration is compensated by increases in concentrations of base cations from the exchange sites. In surface waters, ion exchange is insignificant, and alkalinity is independent of P_{CO_2} ; here most of the response to change in P_{CO_2} is in pH (Fig. 3).

Natural organic acids play a major role in podzolic soils. Organic anions can comprise a significant fraction of the anion sum in soil solution in surface horizons^{29,30}. These organic compounds, however, are typically precipitated as organo-aluminium complexes in deeper soil horizons such as the Bh or Bhs horizons in spodosols³⁰. This process plays a major role in the formation of the spodic layer in the podzolic soils that are widespread in glaciated regions of North America and Europe. Thus although surface soil solutions might contain high concentrations of organic anions and be highly acidic, organic anion concentrations and acidity are lower in subsoil solutions and runoff. At the Adirondack sites in New York, USA, studied by Cronan and Aiken³⁰, for example, dissolved organic carbon concentrations were 21–32 mg C l^{-1} in surface soil leachate, but only 5–7 mg C l^{-1} in B horizon leachates, and 2–8 mg C l^{-1} in surface waters.

Increased sulphate concentration in soil solution resulting from increased acid deposition can produce large decreases in

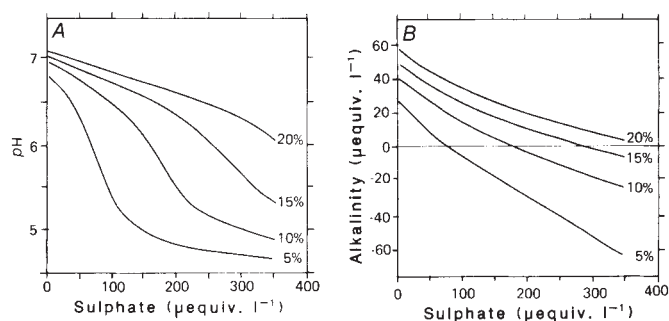


Fig. 5 pH of surface water (A) and soil solution alkalinity (B) for soils in which 5%, 10%, 15%, and 20% of exchange sites are occupied by Ca^{2+} . Calculations were made using the model of Reuss and Johnson¹⁴ assuming P_{CO_2} in soil of $10^{-1.5}$ atm and $10^{-3.5}$ atm for the surface water. Al solubility ($3 pH + \log(\text{Al}^{3+})$) was set at 8.75, and the log of the Al-Ca selectivity coefficient set at 2.25.

soil solution alkalinity (Fig. 4). This decrease in alkalinity results from increased concentrations of acid cations necessary to balance the higher sulphate concentration. The magnitude of decrease in alkalinity depends on the soil exchange and aluminium solubility reactions. Decreases will be larger in acid soils. Surface waters draining such soils are thus highly sensitive to the concentration of strong acid anions (SO_4^{2-}). At low levels of base saturation the alkalinity can decrease from positive to negative and thus result in a major change in the pH of the surface water (Fig. 5).

Together the process of sulphate adsorption, cation exchange, dissolution of soil minerals and dissolution of CO_2 can account for the general equilibrium relationships between acid deposition, soil, soil solution, and surface waters.

Flux relationships

The equilibrium reactions between soil and soil solution address surface water quality at a particular time. Over the long-term, however, the flux of acids and bases into and out of the soil may change the chemical state of the soil thus affecting the soil/soil solution equilibria.

Sulphate adsorption and the base cation status are two key capacity factors. Acid deposition progressively fills sulphate adsorption sites which in turn leads to increased concentrations of sulphate in soil solution and a flux of sulphate through the soil. The increased concentration of anions in soil solution necessitates increased concentrations of cations including base cations. Increasing the flux of sulphate in runoff will thus increase the flux of base cations out of the soil. If not replenished by base cation inputs from the atmosphere or weathering of soil minerals, the long-term response will be a decrease in the pool of base cations on the cation exchange complex.

The rate of change in base saturation is the net sum of fluxes of base cations into and out of the soil:

$$\text{dBCX}/\text{dt} = \text{AD} + \text{W} - \text{B} - \text{Q}(\text{BC}) \quad (2)$$

where dBCX/dt is the rate of change of exchangeable base cations on the soil, AD is the atmospheric deposition of base cations, W is the release of base cations from mineral weathering, B is net incorporation into biomass, and $\text{Q}(\text{BC})$ is the export of dissolved base cations in drainage waters (given by the product of runoff volume and base cation concentration).

For a typical shallow acidic podzol the annual flux of base cations is a few per cent of the total stored in the cation exchange pool³¹ (Fig. 6). Even moderate levels of sulphate deposition with the subsequent increased anion flux through the soil will cause a significant increase in flux of base cations from the exchange sites.

Long-term changes in soils also lead to changes in surface

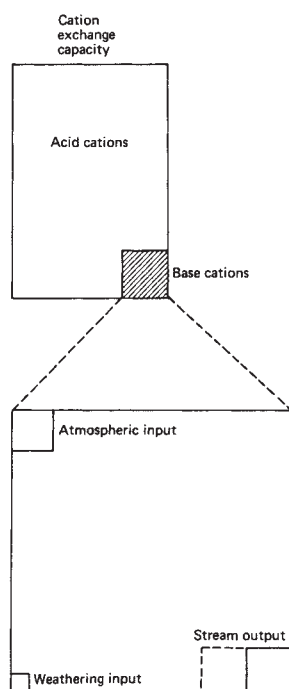


Fig. 6 The relative sizes of cation pools and yearly cation fluxes for a typical soil in a sensitive area receiving acidic deposition. We use here data for an acidified forest lake on the Swedish west coast³¹. The upper box depicts the total exchangeable cation pool in the soil; 95% are acid cations (H^+ , Al^{3+}) and only 5% are base cations (Ca^{2+} , Mg^{2+} , Na^+ , K^+). In the lower box the pool of base cations is enlarged to illustrate pool size in relationship to the annual fluxes of base cations into and out of the soil. Under steady-state conditions the annual flux of base cations leaving the soil in streamwater equals the annual inputs of base cations from the atmosphere and weathering. Acid deposition at Gårdsjön has increased the stream output of base cations (dashed box), and the pool of exchangeable base cations in system is presently decreasing.

water pH and alkalinity (Fig. 7). Here the type and rate of response depends on capacity factors, primarily sulphate adsorption and size of pool of exchangeable base cations, that is, the base saturation.

Mathematical models provide a means by which these equilibrium and flux relationships can be quantitatively coupled to examine the effects of acid deposition on soil and water acidification^{23,32-35}. These models consist essentially of a combination of a set of chemical equilibria equations with a set of book-keeping procedures by which the fluxes of cations and anions into and out of the soil are tallied. Because the chemical reactions are nonlinear, their interactions produce nonlinear and hysteretic behaviour. Mathematical models of these interactions can expose the full range of implications inherent in assuming these chemical processes control soil and surface water response to acidic deposition. Successful application of such numerical models to catchments in many parts of the world illustrates the general utility of this approach. Such process-oriented models offer a promising tool for the prediction of the future acidification status of soil and water³⁶⁻³⁸.

Discussion

Acidification of soils and surface waters due to the deposition of strong acids from the atmosphere is superimposed upon natural acidification processes. Natural acidification is dominated by carbonic and organic acids both of which are products of biological activity in the terrestrial environment. These acids are responsible for the acid podzolic soils characteristic of

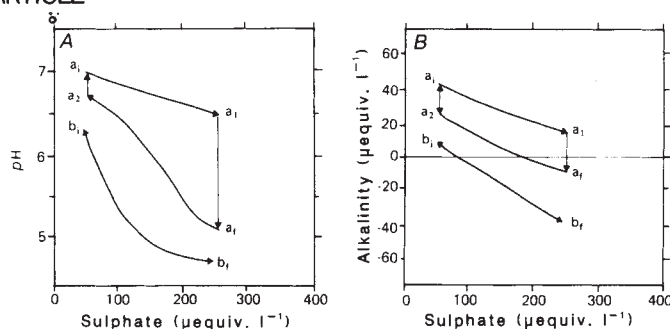


Fig. 7 An illustration of the pathways followed by pH and alkalinity of surface waters in response to changes in sulphate concentrations for two of the soils (20% and 5% base saturation, A and B, respectively) shown in Fig. 5. Pathway a_1 to a_2 is typical of systems with moderately buffered soils (base saturation 20%). Surface water alkalinity and pH decline slightly in response to increased SO_4^{2-} flux through the soils, but the largest changes in alkalinity and pH occur over the longterm as the soils are leached of base cations. Pathway b_1 to b_2 is typical of acid soils (base saturation 5%). Here loss of alkalinity and sharp declines in pH occur immediately as the SO_4^{2-} flux increases, with further reductions coming as the base saturation of the soil is reduced. Recovery of surface water pH and alkalinity following decreases in sulphate deposition can also be inferred. As sulphur deposition declines (and sulphate flux through the soil decreases) the system will move along lines of constant base saturation (a_2 to a_1) until the leaching of base cations accompanying the sulphate flux has been reduced sufficiently such that mineral weathering replenishes base cations on the exchange complex. As base saturation increases (a_2 to a_1), the system returns to original state.

sensitive regions of North America and Europe.

Carbonic acid or the associated bicarbonate processes cannot give rise to acidic surface waters. Large overpressures of P_{CO_2} are necessary to yield pH levels below 5.0 (Fig. 3). Surface waters with P_{CO_2} levels at or near atmospheric will have significant bicarbonate concentrations only if pH is above 5.5. Bicarbonate thus is not the mobile anion responsible for widespread surface water acidification.

Natural organic acids can cause acidic ($pH < 5.0$) surface waters in regions not receiving significant acidic deposition, for example, in bog waters in central Canada³⁹, several lakes in Florida⁴⁰, and rivers in the Amazon region⁴¹. Such waters are usually highly coloured. Two chemical conditions must be met for these waters to be acidic ($pH < 5$); the sum of concentrations of strong acid and organic anions must exceed that of base cations, and the organic acids must be significantly dissociated at $pH < 5.0$ ($pK < 5.0$). High colour or high concentration of dissolved organic carbon (DOC) alone thus does not ensure high acidity.

The effect of acid deposition on soil solution and surface water depends on both intensity and capacity factors. Intensity effects include the changes in soil solution and surface water chemistry that occur in direct response to increases in strong acid anion concentrations brought about by acid deposition. Soil solutions are buffered by a combination of exchange and dissolution processes (Fig. 2). In soil with low base saturation, the H^+ and Al species brought into solution as a result of acid deposition may be sufficient to cause a shift from positive to negative alkalinity, resulting in acidic surface water (Figs 5 and 7). Acidification of surface waters can occur on the short term without first requiring a change in base saturation.

The principal capacity effects are sulphate adsorption and the increased leaching of base cations accompanying the mobile strong acid anions. Such long-term changes in soils have been reported from acid-impacted areas such as Germany and southern Sweden⁴²⁻⁴⁵.

Acid addition experiments in Norway exemplify the decisive

role of atmospheric inputs of strong acids in the acidification of runoff. Beginning in 1984 70–100 kequiv. km⁻² yr⁻¹ of H₂SO₄ has been added to a 0.7 ha pristine headwater catchment. Prior to treatment, runoff draining the thin and patchy acidic soils (pH 4.2–5.8) was of Ca-Mg bicarbonate composition with pH 5.4–6.5, typical of waters in sensitive areas. Acid addition at levels comparable to those in acidified areas of southernmost Norway caused large and rapid changes in runoff chemistry; pH decreased to below 5, sulphate became the dominant anion, and aluminium reached levels toxic to fish^{46,47}.

Changes in land use affect the base status of soils, and uptake of bases by aggrading forests can result in soil acidification through depletion of the exchangeable base cation pool^{14,44,45,48}. The rate of depletion may be very similar to that caused by moderate level of acidic deposition^{49,50}. The transfer of acidity from soils to surface water, however, can only be accomplished by "carrier" anions that are not protonated in acid soil solution. Mobile anions are thus required to produce negative alkalinity in surface waters^{51,52}. That the transfer of acidity may be accomplished by organic anions is attested to by the acidity of some highly coloured waters both in regions receiving acid deposition and in pristine regions^{39,40,53}. In the case of the recently acidified lakes and streams in eastern North America and Europe, however, sulphate is generally the dominant mobile anion^{2,8,9,11,13}.

Krug and Frink¹⁶ have proposed that acidification of lakes and streams entails a switch from organic-acid dominated waters to sulphate dominated waters without a large change in pH. If this is the case, however, then the typical sulphate-rich acidified lake today would have been extremely coloured in the past. Several studies indicate that dissolved organic matter (DOC) contains about 4–11 µequiv. l⁻¹ of organic anions per mg DOC l⁻¹ (refs 30, 54, 55), and thus the typical Adirondack Park lake with present-day sulphate concentration of about 120 µequiv. l⁻¹ would have had an additional 10–30 mg DOC l⁻¹ in the past. There is no evidence for such major change. Furthermore the vast majority of lakes in areas not receiving acid deposition have pH levels above 5.5. Of 720 lakes in the western United States only 1 had pH below 5.0 and it was associated with a mineral hot spring¹⁰. Repeated surveys of lakes in northern Norway encountered no lake with pH below 5.5 (ref. 56), and of 198 lakes in Labrador only 1 had pH below 5.0 (ref. 9). The empirical data simply do not suggest a major role for organic acids in the widespread recent acidification of lakes in eastern North America and northern Europe.

Conclusions

Key soil processes in concert with inputs of strong acids from the atmosphere explain the observed trends in soil and water acidification in North America and Europe. The apparent paradox of waters with pH > 5.5 draining acidic soils in pristine areas not receiving significant acid deposition is readily understood as the result of the simultaneous effect of inorganic carbon equilibria, cation exchange, inorganic aluminium equilibria and chemical weathering. These are well-known and important components in soil solution chemistry; it is their simultaneous interaction that results in the apparent paradox. The elevated P_{CO₂} in soils results in acidic soil solution with positive alkalinity; upon emerging as streamwater degassing occurs and pH increases. The result is pH > 5.5 waters draining acidic soils found in such areas as Alaska, northern Canada, northern Norway and northern Sweden.

These equilibria and flux relationships also explain a second apparent paradox. Acidic soils contain exchangeable H⁺ and Al³⁺ in amounts equivalent to thousands of years of acid deposition; yet acid deposition to such soils causes acidification of runoff with subsequent loss of fish and other biological effects. In the absence of mobile strong acid anions the enormous store of acid in such soils seldom results in the acidification of freshwaters to the stage at which alkalinity becomes negative;

H⁺ and inorganic Al³⁺ are not mobilized to levels toxic to fish. Addition of relatively small amounts of strong acids to such acid soils, however, can have a major effect on the acidity and concentration of inorganic Al in runoff. Mobile strong acid anions such as SO₄²⁻ and NO₃⁻ readily transport H⁺ and Al³⁺ out of the soil. Whereas degassing will reduce bicarbonate anion concentrations, these strong acid anions are unaffected. The result is the acidic, aluminium-rich sulphate waters found in sensitive regions receiving acid deposition such as southern Quebec, Adirondack Mountains of New York, southwestern Scotland, southern Norway and southern Sweden.

We thank D. F. Brakke and S. A. Norton for helpful comments. Our studies evolved from W. Dickson's statement in 1980 "Show me your anion and I will tell you who you are". This work was supported in part by the Royal Norwegian Council for Scientific and Industrial Research, the Surface Water Acidification Programme (SWAP—The Royal Society, The Norwegian Academy of Science and Letters, and The Royal Swedish Academy of Sciences), the Colorado Agriculture Experiment Station, and the US Environmental Protection Agency (EPA).

1. Wright, R. F. & Henriksen, A. in *International Symposium on Sulphur Emissions and the Environment* 277–301 (Society of Chemical Industry, London, 1979).
2. Wright, R. F. *Water Qual. Bull.* **8**, 137–142 (1983).
3. Haines, T. A. in *Acid Deposition Long-Term Trends* 300–334 (National Academy Press, Washington, DC, 1986).
4. Galloway, J. N., Likens, G. E. & Hawley, M. W. *Science* **226**, 829–831 (1985).
5. Likens, G. E. *Environ. Sci. Tech.* **10**, 29A–44A (1976).
6. Galloway, J. N. & Cowling, E. B. *J. Air Pollut. Cont. Assoc.* **28**, 229–235 (1978).
7. Wright, R. F. & Gjessing, E. T. *Ambio* **5**, 219–223 (1976).
8. *Monitor 1986 Acid and Acidified Waters* (Swedish National Environmental Protection Board, Solna, Sweden, 1986) (in Swedish).
9. Jeffries, D. S., Wales, D. L., Kelso, J. R. M. & Linthurst, R. A. *Water Air Soil Pollut.* **31**, 551–568 (1986).
10. Eilers, J. M., Brakke, D. F. & Landers, D. H. *Verh. Internat. Verein. Limnol.* (in the press).
11. Wright, R. F. & Henriksen, A. *Limnol. Oceanogr.* **23**, 487–498 (1978).
12. Dupont, J. & Grimard, Y. *Water Air Soil Pollut.* **31**, 223–230 (1986).
13. Linthurst, R. A. *et al. Water Air Soil Pollut.* **31**, 577–592 (1986).
14. Reuss, J. O. & Johnson, D. W. *Acid Deposition and the Acidification of Soils and Waters*. (Springer, New York, 1986).
15. Rosenqvist, I. T. *Sci. Total Environ.* **10**, 39–49 (1978).
16. Krug, E. C. & Frink, C. R. *Science* **221**, 520–525 (1983).
17. Van Breeman, N. *et al. Nature* **229**, 548–550 (1982).
18. Wright, R. F. *Predicting Acidification of North American Lakes* (Acid Rain Res. Rept. 4/83, Norwegian Inst. Water Res., Oslo, 1983).
19. Kramer, J. R. *et al. in Acid Deposition Long-Term Trends*, 231–299 (National Academy of Sciences, Washington, DC, 1986).
20. Johnson, D. W. & Cole, D. W. *Environ. Internat.* **3**, 79–90 (1980).
21. Chao, T. T., Harward, M. E. & Fang, S. C. *Soil Sci. Soc. Am. Proc.* **26**, 234–237 (1962).
22. Johnson, D. W. & Todd, D. E. *Soil Sci. Soc. Am. J.* **47**, 792–800 (1983).
23. Cosby, B. J., Hornberger, G. M., Galloway, J. N. & Wright, R. F. *Water Resour. Res.* **21**, 51–63 (1985).
24. Johnson, D. W. & Henderson, G. S. *Soil Sci.* **128**, 34–40 (1979).
25. Smith, R. A. & Alexander, R. B. *Nature* **322**, 722–724 (1986).
26. Lindsay, W. L. *Chemical Equilibria in Soils* (Wiley, New York, 1979).
27. Coleman, N. T. & Thomas, G. W. in *Soil Acidity and Liming*, (eds Pearson, R. W. & Adams, F.) (American Society for Agronomy, Madison, 1967).
28. Stumm, W. & Morgan, J. J. *Aquatic Chemistry* (Wiley, New York, 1981).
29. Johnson, D. W., Cole, D. W., Gessel, S. P., Singer, M. J. & Minden, R. V. *Arctic Alpine Res.* **9**, 329–343 (1977).
30. Cronan, C. S. & Aiken, G. R. *Geochim. cosmochim. Acta* **49**, 1697–1705 (1985).
31. Andersson, F. & Olsson, B. (eds) *Lake Gårdsjön, An Acid Forest Lake and Its Catchment* (*Ecol. Bull. (Stock.)* **37**) (1985).
32. Reuss, J. O. *Ecol. Modelling* **11**, 15–38 (1980).
33. Christophersen, N., Seip, H. M. & Wright, R. F. *Water Resour. Res.* **18**, 977–997 (1982).
34. Schnoor, J. L., Palmer, W. D. Jr & Glass, G. E. in *Modelling of Total Acid Precipitation Inputs* (ed. Schnoor, J. L.) 155–174 (Butterworth, Stoneham, Massachusetts, 1984).
35. Chen, C. W., Gherini, S. A., Dean, J. D., Hudson, R. J. M. & Goldstein, R. A. in *Modelling of Total Acid Precipitation Inputs*, (ed. Schnoor, J. L.) 175–204 (Butterworth, Stoneham, Massachusetts, 1984).
36. Gherini, S. A., Mok, L., Hudson, R. J. M., Davis, G. F., Chen, C. W. & Goldstein, R. A. *Water Air Soil Pollut.* **26**, 425–459 (1985).
37. Cosby, B. J., Hornberger, G. M., Galloway, J. N. & Wright, R. F. *Environ. Sci. Tech.* **19**, 1144–1149 (1985).
38. Wright, R. F. & Cosby, B. J. *Atmos. Environ.* **21**, 727–730 (1987).
39. Gorham, E., Eisenreich, S. J., Ford, J. & Santelmann, M. V. in *Chemical Processes in Lakes* (ed. Stumm, W.) 339–363 (Wiley, New York, 1985).
40. Brakke, D. F., Eiler, J. M. & Landers, D. M. in *Proc. UNESCO/IHP-III Symp. Acidification and Water Pathways* Vol. 1, 281–290 (Norwegian Hydrologic Committee, Oslo, 1987).
41. Gibbs, G. E. *Geochim. cosmochim. Acta* **36**, 1061–1066 (1972).
42. Matzner, E. & Ulrich, B. in *Effects of Acidic Deposition on Forests, Wetlands and Agricultural Ecosystems*, (ed. Hutchinson, T. C.) (Reidel, Dordrecht, in the press).
43. Tyler, G., Berggren, D., Bergkvist, B. & Falkengren-Grerup, U. in *Effects of Acidic Deposition on Forests, Wetlands and Agricultural Ecosystems* (ed. Hutchinson, T. C.) (Reidel, Dordrecht, in the press).
44. Tamm, C. O. & Hallbäck, L. *Water Air Soil Pollut.* **31**, 337–342 (1986).
45. Falkengren-Grerup, U. *Oecologia* **70**, 339–347 (1986).
46. Wright, R. F. *et al. Water Air Soil Pollut.* **30**, 47–64 (1986).

47. Wright, R. F. & Gjessing, E. *RAIN project. Annual report for 1985*. (Acid Rain Res. Rept. 10/86, Norwegian Inst. Water Res., Oslo, 1986).
48. Nilsson, L., Miller, H. G., & Miller, J. D. *Oikos* **39**, 40–49 (1982).
49. Johnson, D. W., Richter, D. D., Lovett, G. M. & Lindberg, S. E. *Can. J. Forest Res.* **15**, 772–782 (1985).
50. Richter, D. D., Johnson, D. W. & Todd, D. E. *J. Environ. Qual.* **14**, 263–270 (1983).
51. Gorham, E., *Nature* **181**, 106 (1958).
52. Seip, H. M., in *Ecological Effects of Acid Precipitation*, (eds Drablos, D. & Tollan, A.) 358–366 (SNSF-project, 1432 Ås, Norway).
53. Brakke, D. F., Henriksen, A. & Norton, S. A. *Nature* (in the press).
54. Oliver, B. G., Thurman, E. M. & Malcolm, R. L. *Geochim. cosmochim. Acta* **47**, 2031–2035 (1983).
55. Henriksen, A. & Seip, H. M. *Water Res.* **14**, 809–813 (1980).
56. Wright, R. F., Dale, T., Henriksen, A., Gjessing, E., Johannessen, M., Lysholm, C. & Støren, E. *Regional Surveys of Small Norwegian Lakes* (Report IR 33/77, SNSF-project, 1432 Ås, Norway, 1977).

ARTICLES

Viscous fingering on percolation clusters

Unni Oxaal*, Michael Murat†, Finn Boger*, Amnon Aharony†, Jens Feder* & Torstein Jøssang*

* Institute of Physics, University of Oslo, Box 1048, Blindern, 0316 Oslo 3, Norway

† School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Israel

The displacement of a high-viscosity fluid by a low-viscosity fluid leads to fractal viscous fingering in homogeneous porous media. A study of the viscous fingering at high and low displacement rates in percolating models that are inhomogeneous on all length scales makes it possible to construct different theoretical models for these two rates. Applying these two models to computer simulations on exactly the same geometry as in the experiments suggests that the underlying geometry determines the viscous fingering structure almost entirely.

THE physical phenomena that occur when a low-viscosity fluid is forced into a high-viscosity one, inside a porous medium—such as water pushing oil in a rock—are clearly of much practical interest. These phenomena became the centre of much recent scientific interest when it was realized¹ that under some conditions flow instabilities yield viscous fingers (VF)^{1–11} that are fractal^{7–11}. Fractals are self-similar objects, which look the same at different magnifications¹². Their number of sites, or mass $M(L)$ within a region of linear scale L behaves as

$$M(L) \sim L^D \quad (1)$$

with a usually non-integer fractal dimensionality, D .

The interest in viscous fingers grew when it was observed that for two immiscible fluids, with a high capillary number for the displaced one, the fractal shapes of the VF (generated experimentally in two-dimensional Hele-Shaw cells)^{7,8,11} and in two-dimensional porous media^{9,10} have strong qualitative and some quantitative similarities to those observed in a diverse range of apparently different problems, including dielectric breakdown¹³, diffusion-limited electrodeposition or growth in aqueous solution^{14–17} and computer simulations of diffusion-limited aggregation (DLA)¹⁸. All of these structures have fractal dimensionalities $D = 1.70 \pm 0.05$ in two dimensions. These striking similarities suggested the existence of universality among growth patterns. This suggestion was put forward by Paterson¹ who showed that two-fluid displacement in porous media can be described by Laplace's equation, with boundary conditions similar to those used in the DLA problem. As DLA is a process that is simple to simulate numerically (particles are released at the boundary, and perform random walks until they hit (and stick to) the growing aggregate¹⁸), it was suggested that such simulations may be used to imitate two-fluid flow under various conditions^{1,19}. Recent modifications of the DLA model also allow a quantitative simulation of the time dependence of finger growth in porous media²⁰.

The expectation that flow in Hele-Shaw cells (two parallel glass plates, with a small separation) exhibits the same properties as in real porous media is based on Darcy's law, relating the fluid velocity v to the pressure gradient in the viscous fluid being

displaced,

$$v = -\frac{k}{\mu} \nabla P \quad (2)$$

where k is the average permeability of the porous medium and μ is the fluid viscosity. For incompressible fluids, $\nabla \cdot v = 0$ and therefore,

$$\nabla \cdot (k \nabla P) = 0 \quad (3)$$

If the permeability k is constant over the whole system (as it is in conventional Hele-Shaw cells^{2–4,7}), equation (3) reduces to the Laplace equation, $\nabla^2 P = 0$. If the viscosity of the 'pushing' fluid is negligible, then the pressure in it is uniform, $P = P_0$. Neglecting the surface tension, this is also the pressure in the displaced fluid, at the interface. The pressure on the free boundaries of this fluid is equal to another constant, $P = P_{\text{ext}}$. As the same Laplace equation, with similar boundary conditions, applies to the electrostatic potential in the dielectric breakdown problem¹³, and to the equilibrium probability density of random walkers in the DLA problem¹⁸, the analogy between all the problems follows¹.

In real porous media, the local permeability k fluctuates randomly in space. Therefore, the arguments connecting DLA and VF must be critically reconsidered. Chen and Wilkinson⁹ solved the flow equation (2) inside a square-lattice network of tubes with a statistical distribution of radii using a deterministic growth algorithm. Their resulting viscous fingers were similar to DLA clusters, with $D \approx 1.72$.

Although Chen and Wilkinson⁹ introduced a distribution of channel radii, they had no tubes which were completely blocked. In real porous media, there is usually a finite concentration, $(1-p)$, of blocked paths (which may either be completely blocked or be so narrow that no fluid flows in them in practice). As the concentration p decreases, approaching the percolation threshold p_c , there appears a correlation length $\xi \sim |p - p_c|^{-\nu}$, with $\nu = \frac{4}{3}$, such that the percolating network of open pores has a fractal geometry on length scales²¹ $L < \xi$. In many cases, physical phenomena exhibit a sharp crossover from a fractal behaviour $L < \xi$ to that of a homogeneous random system $L > \xi$ (refs 22–24). Although the work of Chen and Wilkinson⁹ demon-

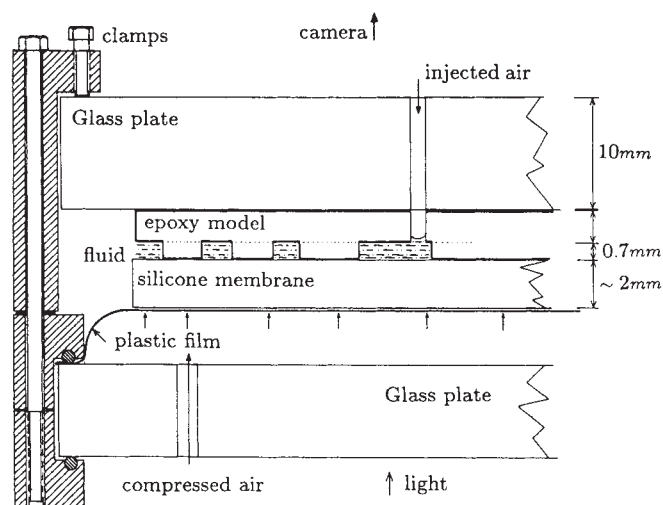


Fig. 1 The experimental setup. The model consists of an epoxy cast with the pore pattern on one side that is closed by a silicone rubber sheet. This assembly is supported by glass plates. The epoxy and the silicone rubber sheets are kept in contact by pressurized air as shown. This setup is transparent, lighted from below and photographed from above.

strated that VF have the same behaviour for the ($p = 1$) or $L > \xi$ case as is found in the uniform Hele-Shaw cells²⁵⁻²⁷ it has not yielded results for the fractal regime, $L < \xi$. This regime was recently studied²⁸, using computer simulations of a stochastic growth model similar to that used for dielectric breakdown¹³. Viscous fingers on the fractal network which occurs at p_c turned out to be very different from those of the homogeneous regime. In particular, their fractal dimension decreased to ~ 1.3 .

As discussed below the computer simulations involve a variety of approximations and of *ad hoc* procedures. It is therefore not at all clear that they represent the true physical situation. On the other hand, real porous media contain too many uncontrolled parameters, that make it difficult to compare them with a theoretical analysis. In order to study viscous fingers on a controlled fractal medium, we performed both real experiments and computer experiments on the same model two-dimensional dilute system, at the percolation threshold.

Geometry

Although the above discussion related to blocked bonds, or tubes, percolation phenomena have the same behaviour for site or bond dilution²²⁻²⁴. For our simulations and experiments we chose a specific realization of a site-diluted square lattice, of size 147×147 , at the site percolation threshold $p_c = 0.593$, in which the central site belongs to the spanning cluster. Flow was allowed only between neighbouring occupied sites, and was possible only on the spanning cluster which connects the site at the centre of the lattice (chosen to be on this cluster) with the boundaries. Figure 2 shows a picture of this cluster, where all the finite clusters were eliminated.

In the basic process to be analysed, the system is first filled with an incompressible viscous fluid (glycerol dyed with Negrosine, in our experiments), and then a non-viscous fluid (air in our case) is pushed at a fixed rate into the central site, pushing the other fluid out at the outer boundaries. As the displaced fluid is incompressible, the viscous fingers cannot move into dangling branches of the cluster, which are connected to the rest of the percolating cluster via a single site. Therefore, the flow is limited to the backbone of the cluster. The backbone of our cluster is also shown in Fig. 2, in white.

In two dimensions, the percolating cluster at p_c has a fractal dimension of $D = 91/48$ (ref. 29), while the backbone has a fractal dimension of $D_{BB} \approx 1.61$ (refs 30, 31). In our 147×147 sample, their respective masses were 6,261 and 3,341 sites.

Experimental setup

We constructed a porous model with a pore structure given by the percolation cluster in Fig. 2, using a method developed by Bonnet and Lenormand³². The model consists of an epoxy resin plate where the pores and connecting channels form a recessed pattern. The epoxy resin plate was made in the following way. A plotter was used to draw the desired pattern on a drafting sheet, and a photographic transparency of the proper size was made from this original. The channels appear in black while the regions where flow will not be permitted are transparent. The film was placed onto a photosensitive nylon plate (BASF Nyloprint WA 175), and this assembly was exposed to ultraviolet radiation. This hardened the unmasked regions of the nylon plate. The masked regions remained soft and were subsequently etched with water, leaving the desired pattern of channels in the nylon plate.

The nylon model is not suitable for the experiments, mainly because the nylon absorbs fluid and swells, distorting the geometry. Therefore we made a silicone rubber mould (RTV 700 and catalyst Beta 4 from General Electric) from the nylon plate, and cast the final epoxy (Araldite XW 396 with catalyst XW 397) model using this mould.

The model is closed by a 2-mm flexible silicone rubber membrane (Eccosil 2CN with catalyst 50 from Emerson and Cuming) held in contact with the plate by compressed air. This restricts the fluid flow to the channels. Reservoirs were made at the boundaries to keep the model open to the flow and the boundary pressure at ambient conditions. The dimensions of the model are 16.5×16.5 cm, the pore volume is $\approx 1 \mu\text{l}$, channel depth 0.7 mm, pore diameter 1.1 mm and bond width 0.7 mm. To ensure uniform wettability the model and membrane are coated by absorbing a monolayer of collagen, from a 2% solution. Water and glycerol wet models treated in this way in a uniform manner.

The model is filled with glycerol coloured black by water-soluble Negrosine, before it is closed. The model is supported from both sides by 10-mm glass plates. The lower of the two glass plates is clamped together with a plastic film, and pressurized air is applied to the space between. In this way we force the silicone rubber membrane into contact with all the regions of the model where fluid flow is prohibited. The air pressure was kept constant and sufficiently high to keep the membrane in place during the experiment.

In the experiments the glycerol in the channels was displaced by air, injected through a 0.7-mm hole in the epoxy model. The point of injection was chosen at the cluster site at the geometrical centre of the model. The air was injected at constant pump rate, using a piston type pump (Pharmacia P-500). The displacement process was photographed from above using a Nikon F3 camera at time intervals controlled by an IBM PC. Lighting was provided from below. The experimental setup is shown in Fig. 1.

Initially the network of channels appears black on a transparent background. Pores invaded by air appear white, as shown in Fig. 3a. As shown in the insert, some glycerol remains on the walls of the channels due to wetting effects.

Experimental results

The experiments were performed at several flow rates. At low flow rates a typical interval between photographs was 30 s while at high flow rates it was typically 1.3 s. Basically, we find all the experiments with rates larger than $\sim 50 \text{ ml h}^{-1}$ to have very similar results. Figure 3a shows a typical picture of such a 'fast' experiment, for a flow rate of 100 ml h^{-1} , 29 s after the start of the experiment.

To analyse the results quantitatively we measured the coordinates of the invaded pores on the photographs using a Tektronix 495 digitizing tablet controlled by an IBM PC. Figure 3b shows the same data as in Fig. 3a, and Fig. 4a shows the result of a 'slow' experiment, at a flow rate of $1 \text{ ml h}^{-1} \sim 0.33 \text{ pores s}^{-1}$.

Both digitized pictures contain four sets of data, taken at different times in the same experiment, with the masses of the growing aggregate as indicated.

The fast and slow experiments can be quantitatively distinguished by looking at plots of the mass $M(t)$ of the fingers (the volume of injected air), counted on the digitized pictures, as a function of time, t . As can be seen from Fig. 5, there is a qualitative difference between the functions $M(t)$ for the two kinds of experiments. Figure 5a shows $M(t)$ for the slow flow. As our experiments were performed using a constant pump rate when injecting the air, the linearity of Fig. 5a (with a measured slope equal to the pump rate), proves that the motion of the interface in the slow flow does not depend on the pressure gradient in the displaced fluid. This is inconsistent with equation (2), and with the stochastic DLA (or dielectric breakdown) algorithm. This result led us to apply a new simulation model, described below.

At high injection rates, Fig. 5b shows a nonlinear dependence of $M(t)$. We attribute this nonlinearity to the increase in the pressure gradient in the viscous fluid, as the interface approaches the external boundary of the sample. This is consistent with equation (2) and with the stochastic model. An approximate form for the shape of $M(t)$ is obtained as follows: as the cluster is at p_c , there exist several independent channels from the origin to the boundary. Each of these channels consists of singly connected sites, through which air must pass, connecting intermediate blobs³³. The flows in these channels are decoupled from each other. Approximating each of these by a quasi-one-dimensional channel, of total length L (between origin and external boundary), the location x of the interface should obey $dx/dt \sim \nabla P \sim P_0/(L-x)$, hence $t = Ax - Bx^2$. In our experiments, we found that the radial distance from the origin to the end of the fastest tip is proportional to the radius of gyration, $r \sim m^{1/D}$, where m is the mass of the finger. Assuming that $x \sim r$, we expect

$$t = t_0 + a m^{1/D} - b m^{2/D} \quad (4)$$

Here t_0 is the time at the start of the experiment. The full curve in Fig. 5b is a fit to this expression, fixing $1/D = 0.75$. Note that the last two points are below the fitted line, probably due to boundary effects.

Simulations of fast flows

As demonstrated by our discussion of equation (4), we believe that the flow of the viscous fluid is governed by equations (2) and (3), which could in principle be solved explicitly and deterministically⁹. In the case of the fast flow rate, we believe that this will give the same results as the stochastic approach, invented in ref. 13 and used in ref. 28: at every time step, we first solve a discrete version of the Laplace equation inside the viscous fluid,

$$P_{\mathbf{r}} = \frac{1}{4} \sum_{\delta} P_{\mathbf{r}+\delta} \quad (5)$$

where sites $\mathbf{r} + \delta$ are the neighbours of site $\mathbf{r}$. If a site at $\mathbf{r} + \delta$ is blocked, we set $P_{\mathbf{r}+\delta} = P_{\mathbf{r}}$ in equation (5), to ensure the boundary condition $\nabla P = 0$ on the bond from $\mathbf{r}$ to $\mathbf{r} + \delta$. Equation (5) is solved with the pressure equal to 1 on the interface between the fluids, and equal to 0 on the external boundaries. After convergence, a new open perimeter site is added to the displacing fluid region (and assigned $P = 1$), using a stochastic probability which is proportional to $|\nabla P|$ on the bond connecting that site to the growing aggregate^{13,25-27}. We believe that this procedure should imitate the growth, since in our geometry growth occurs only at relatively few tips, and it should not matter if these grow simultaneously (as they would in a deterministic solution) or at randomly alternating time steps.

Figure 3c shows a sequence of growth steps in a typical computer simulation, chosen at the same aggregate masses as in Fig. 3b. Clearly, the results of our simulations are qualitatively

Fig. 2 (Opposite) A spanning percolation cluster on a lattice of 147×147 bonds. The cluster shown consists of 6,261 sites. The concentration of occupied sites is $p_c = 0.593$, but only sites on the spanning cluster are shown. Occupied neighbouring sites are connected by conducting bonds. The backbone of the spanning percolation cluster is shown in white, and consists of 3,341 sites, the remaining sites are light blue.

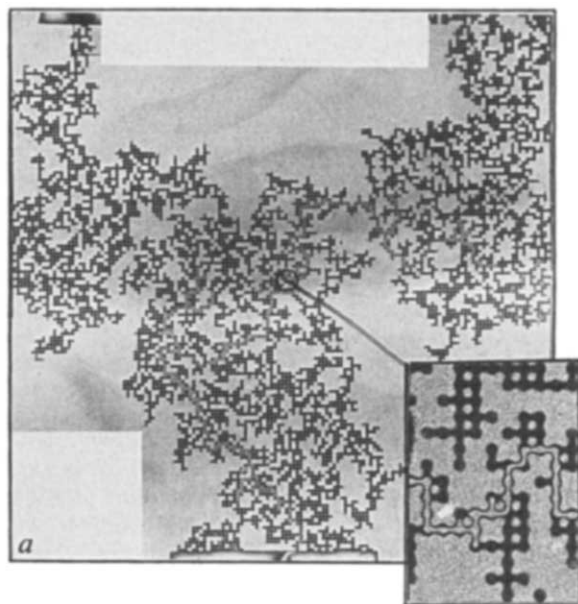


Fig. 3 *a*, Photograph from a 'fast' displacement experiment at $100 \text{ ml h}^{-1} \sim 30$ pores per second, just after breakthrough. Pores filled with glycerol appear black. Pores where the glycerol was displaced by air appear white. The total number of pores invaded by air is 447. As shown in the insert glycerol was trapped in corners of the pores invaded by air because of capillary effects. *b*, A digitized version of the 'fast' experiment shown in *a*. The backbone is black. The sites added during the time interval between two consecutive pictures are coloured differently. The masses are 30, 86, 213 and 447. The centre of injection is marked by a light blue point. *c*, A sequence of growth stages in a typical computer simulation at the same aggregate masses as in *b*. *d*, The 'fast' experiment of *b*, and the simulations from *c*, both at the same cluster mass of 213, superimposed. The experimental result is yellow, the result of a simulation is red. The overlap—sites found both in the simulation and in the experiment—is white.

Fig. 4 (Bottom row, opposite) *a*, A digitized version of the 'slow' experiment at $1 \text{ ml h}^{-1} \sim 0.33$ pore s^{-1} . Colour code as in Fig. 3. The masses are 145, 379, 543 and 753. *b*, Simulation of 'slow' flow with $\eta \rightarrow 0$ at the same aggregate masses as in *a*.

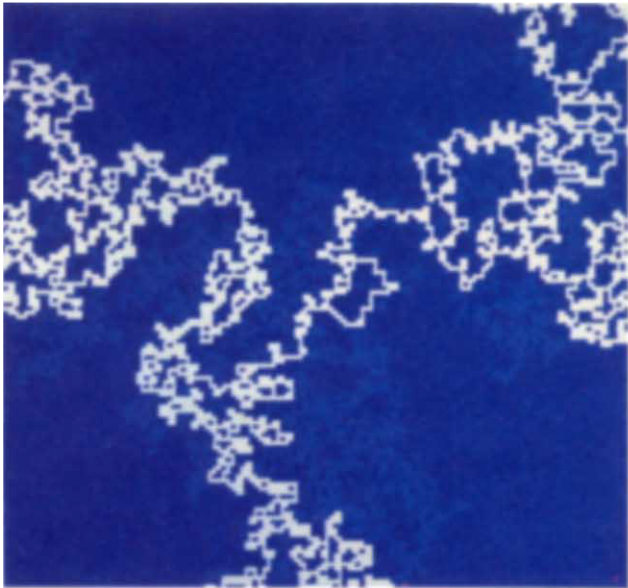


Fig. 2

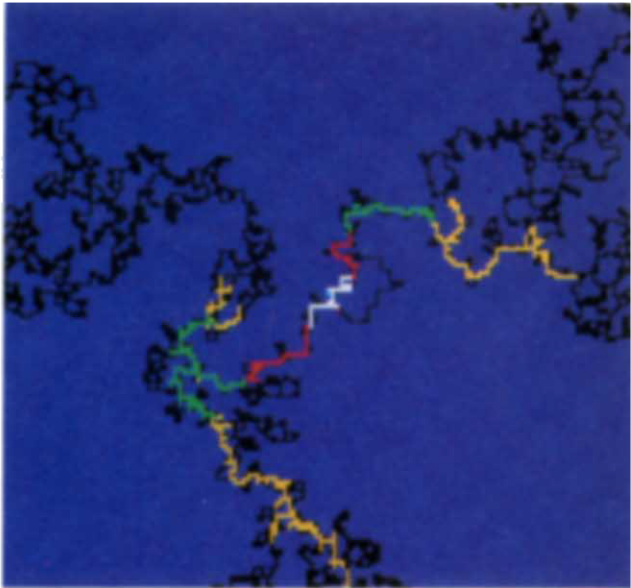


Fig. 3b

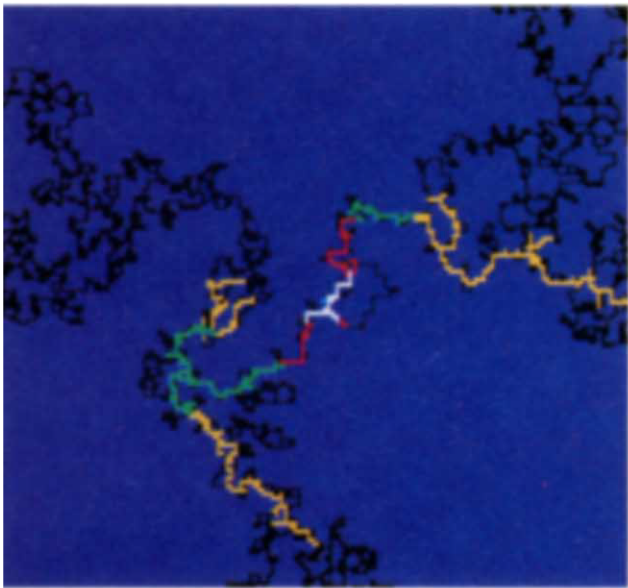


Fig. 3c

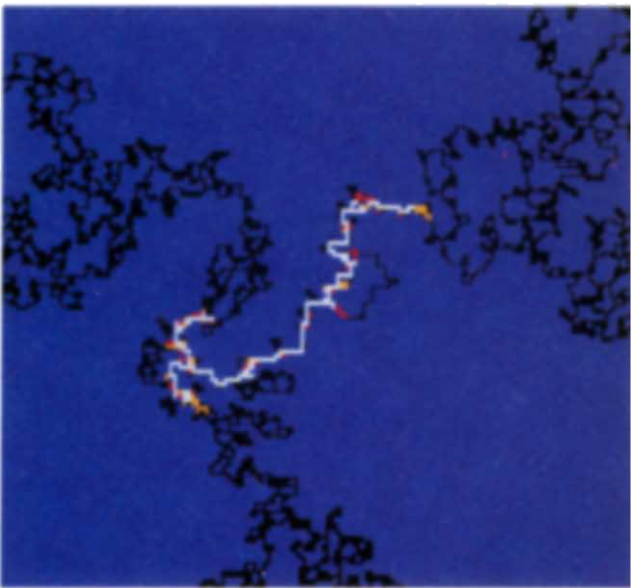


Fig. 3d

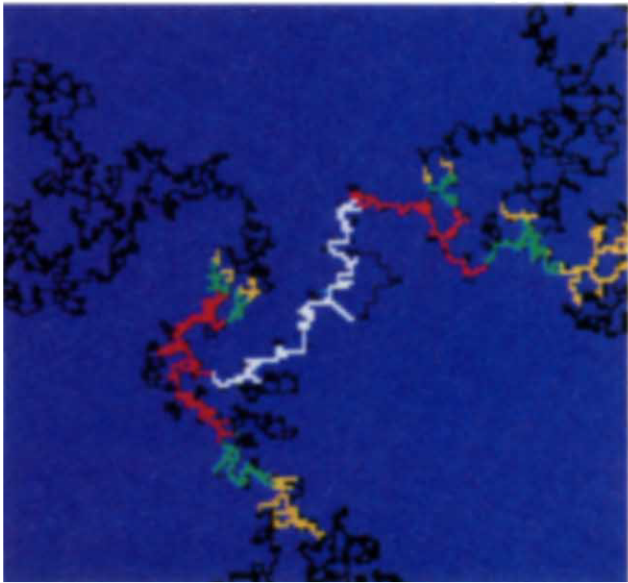


Fig. 4a

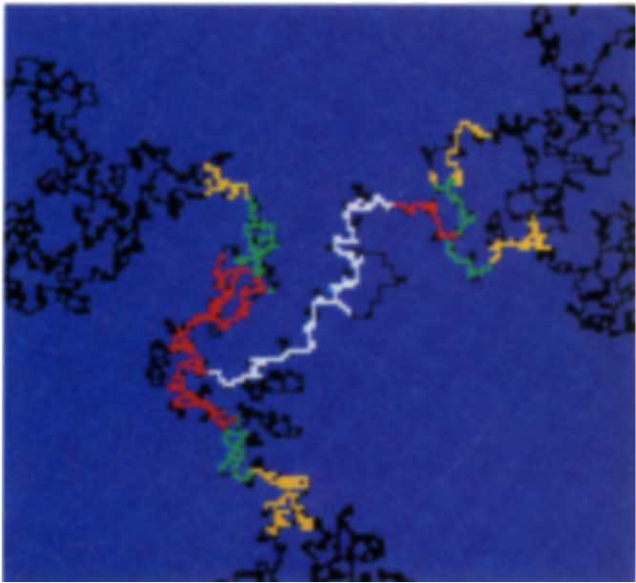


Fig. 4b

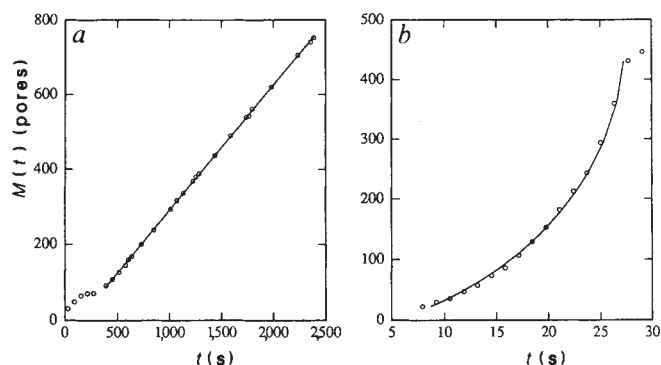


Fig. 5 The mass of the fingers as function of time. *a*, Slow flow, $1 \text{ ml h}^{-1} \sim 0.33 \text{ pore s}^{-1}$ (full line). *b*, Fast flow, $100 \text{ ml h}^{-1} \sim 30 \text{ pores s}^{-1}$. The full line represents the fit of equation (4) to the observations. The parameters in equation (4) are: $t_0 = 3.51$, $a = 0.51$ and $b = 2.9 \times 10^{-3}$.

very similar to the experimental results shown in Fig. 3*b*. To make this more quantitative, Fig. 3*d* shows the two sets of pictures superimposed on each other, when the mass was 213. For these three times (aggregate masses of 30, 86 and 213), the number of overlapping sites are 60%, 86% and 77% of the masses.

To appreciate these amounts of overlap, we repeated the numerical simulation three different times, with different random seeds for the random number generator responsible for the random stochastic choices in the growth procedure. Typical overlaps between two different simulations at the same times as above are 67%, 75% and 79%, comparable to the overlaps between the simulation and the real experiments. Similar overlaps were also found between each of the three simulations and the experiment shown in Fig. 3*b*—the average overlaps were 66%, 80% and 78% with errors of order $\pm 6\%$.

These high values of overlap show that the geometry of the growing fingers is determined to a large extent by that of the underlying backbone. In fact, a finger which connects two sites on the backbone must pass through all the singly connected sites on the link between these two points³³. It is now widely accepted that the numbers of these also grows with the distance L between the two end points, as³⁴ $L^{1/\nu}$. In fact, the actual overlap between different experiments is much larger than this, because the finger also usually picks the shortest path through 'blobs' on the route.

The fact that percolation clusters have a finite order of ramification^{35,36} implies that there is always a small finite number of sites which must be cut to isolate an arbitrarily large region of the cluster, independent of the length scale¹². Apparently, this determines the paths of the fingers, and—through that—their detailed geometry. The fact that the geometry of the particular cluster dominates the growth is also reflected in the dependence of the fingers' mass on the radius of gyration. Figure 6*b* shows points from the three simulations, as well as from the experiment. Although on average all are consistent with the fractal dimension of 1.3 (full line), as was found in ref. 28, all the data show systematic oscillations about the average, reflecting the particular geometry of our specific cluster. Similar lacunarity effects were discussed in ref. 21. Figure 6*a* shows the same phenomenon (with 10 simulations) for a slow flow experiment.

Simulations of slow flows

As demonstrated by Fig. 5*a*, the growth of the fingers in the slow experiments does not depend on the pressure gradients in the viscous fluid. Instead, it is determined by local interfacial forces, rather than viscous ones. If one assumes that all the

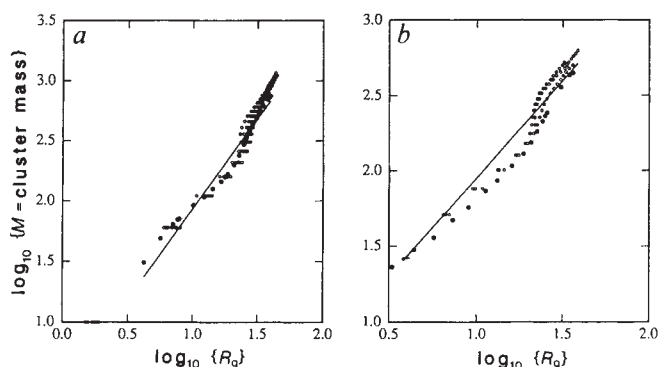


Fig. 6 The mass of the fingers as function of the radius of gyration. *a*, Slow flow, 1 ml h^{-1} ; *b*, fast flow, 100 ml h^{-1} . The slopes of the straight lines are $D = 1.5$ and $D = 1.3$, respectively. Open circles show results of simulations, filled circles show experimental data.

channels are identical, then one would expect equal growth probabilities at all the perimeter sites which allow flow, excluding those which surround isolated lakes of the invaded fluid. As these allowed sites would have finite pressure gradients in the solution of Laplace's equation, $|\nabla P| \neq 0$, we can identify them by associating them with a growth probability

$$p_i \sim |\nabla P|^\eta, \text{ with } \eta \rightarrow 0 \quad (6)$$

where the parameter η was introduced in ref. 13.

For homogeneous systems, the limit $\eta \rightarrow 0$ is usually associated³⁷ with the Eden growth model³⁸, in which all perimeter sites have the same growth probabilities. Indeed, when we simulated the model of equation (6) on a uniform lattice, we found compact aggregates, whose mass scales as the euclidean area, $M \sim L^2$.

This equivalence, between the $\eta \rightarrow 0$ model and the Eden model, breaks down on the percolation cluster. Clearly, the Eden model could fill the whole cluster ($D \sim 1.9$), while the $\eta \rightarrow 0$ model is restricted to the backbone, and hence $D \leq 1.61$. In practice, the incompressibility of the displaced fluid creates many impenetrable regions on the backbone, and D is significantly lower.

Our simulation ran as follows: we start with a seed of the invading fluid at the centre. We then choose at random a perimeter bond of that seed, and add the site on the other side of the bond to the growing aggregate (invading fluid). We then identify all the 'finite' clusters of the invaded fluid, and mark their sites so that the bonds connecting them to the aggregate will never be chosen again. We continue in this manner until a boundary is reached.

Figure 4*b* shows three stages of a simulation of this model on our cluster. Indeed, the aggregate is far from filling the backbone. In fact, averaging over several simulations on 35 different clusters of size 251×251 we estimate its fractal dimension to be ~ 1.5 . Figure 6*b* shows that, apart from the lacunarity effects, both our simulations and our experiments are consistent with this estimate.

As before, we find a large overlap between different runs of the simulation on the same cluster. At aggregate masses of 145, 379 and 543, typical overlaps are 78%, 76% and 80%. It is also rewarding to compare our simulations, Fig. 4*b*, with the slow rate experiments, Fig. 4*a*. The corresponding overlaps are 80%, 80% and 74%, all $\pm 3\%$.

Although the masses of the 'slow' fingers are somewhat higher than those of the 'fast' ones, we note that the geometry of Fig. 4*a* and 4*b* is also dominated by the underlying cluster geometry; in particular it is constrained by the singly connected sites and the finite order of ramification.

Discussion

Our main results may be summarized as follows: (1) the geometry of the percolation cluster, and in particular the finite order of ramification and the singly connected paths, determines the geometry of the viscous fingers on the cluster. This leads to large overlaps between experiments and simulation. (2) Fast flow experiments are dominated by viscous effects, and obey a quasi-one-dimensional Laplace equation. This is characterized by a nonlinear $M(t)$ curve, Fig. 5b. The experiments are reproduced by the simulations with $\eta = 1$. (3) Slow flow experiments are dominated by local capillary effects, and are identified by a linear $M(t)$ curve, Fig. 5a. They are reproduced by simulations with $\eta = 0$. However, unlike in homogeneous networks, these differ from the Eden model, which ignores the incompressibility of trapped fluid regions. (4) Results on individual clusters deviate systematically from the average fractal mass relation $M \sim L^D$, reflecting the lacunarity of the specific clusters.

Here we looked only at the extreme case of percolation clusters at the percolation threshold. As demonstrated in ref. 28, subtle crossover phenomena are expected as the concentration p increases above p_c . These remain for future studies.

Received 15 April; accepted 17 July 1987.

1. Paterson, L. *Phys. Rev. Lett.* **52**, 1621-1624 (1984).
2. Saffman, P. G. & Taylor, G. I. *Proc. R. Soc. A* **245**, 312-329 (1958).
3. Paterson, L. *J. Fluid Mech.* **113**, 513-529 (1981).
4. Bensimon D., Kadanoff, L. P., Liang, S., Sraiman, B. I. & Tang, C. *Rev. mod. Phys.* **58**, 977-999 (1986).
5. Paterson, L., Hornof, V. & Neale, G. *Powder Technol.* **33**, 265-268 (1982).
6. Lenormand, R. & Zaccaro, C. *Phys. Chem. Hydro.* **6**, 497-506 (1985).
7. Nittmann, J., Daccord, G. & Stanley, H. E. *Nature* **314**, 141-144 (1985).
8. Ben-Jacob, et al. *Phys. Rev. Lett.* **55**, 1315-1318 (1985).
9. Chen, J. & Wilkinson, D. *Phys. Rev. Lett.* **55**, 1892-1895 (1985).
10. Måløy, K. J., Feder, J. & Jøssang, T. *Phys. Rev. Lett.* **55**, 2688-2691 (1985).
11. Daccord, G., Nittmann, J. & Stanley, H. E. *Phys. Rev. Lett.* **56**, 336-339 (1986).
12. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
13. Niemeyer, L., Pietronero, L. & Wiesmann, H. J. *Phys. Rev. Lett.* **52**, 1033-1036 (1984).
14. Matsushita, M., Sano, M., Hayakawa, Y., Honjo, H. & Sawada, Y. *Phys. Rev. Lett.* **53**, 286-289 (1984).
15. Sawada, Y., Dougherty, A. & Gollub, J. P. *Phys. Rev. Lett.* **56**, 1260-1263 (1986).
16. Grier, D., Ben-Jacob, E., Clarke, R. & Sander, L. M. *Phys. Rev. Lett.* **56**, 1264-1267 (1986).
17. Ben-Jacob, E., Deutscher, G., Garik, P., Goldenfeld, N. D. & Lareah, Y. *Phys. Rev. Lett.* **57**, 1903-1906 (1986).
18. Witten, T. A. & Sander, L. M. *Phys. Rev. Lett.* **47**, 1400-1403 (1981).

An alternative crossover situation arises if the 'blocked' bonds are allowed to have a non-zero small permeability, $k_i \ll 1$. As soon as k_i becomes non-zero, new channels open between the spanning cluster and neighboring finite clusters, and the behaviour again crosses over to that of the homogeneous system, as found by Chen and Wilkinson⁹. But our results may still apply when k_i is very small. Again, this crossover is also left for the future.

Apart from the percolation case, our results should apply to all other flows on fractal networks which have a finite order of ramification, that is, have crucial 'hot' spots through which the flow must pass.

In conclusion, percolation clusters at p_c turn out to allow a satisfactory description of viscous fingering on ramified fractals, consistent with both experiments and simulations.

We thank Geir Holm for photographic work. Work at Tel Aviv was supported by grants from the US-Israel Binational Science Foundation, the Israel Academy of Sciences and Humanities and the Israel AEC. Work in Oslo has been supported by VISTA, a research cooperation between the Norwegian Academy of Science and Letters and Den norske stats oljeselskap a.S (STATOIL).

19. Kadanoff, L. P. *J. statist. Phys.* **39**, 267-283 (1985).
20. Måløy, K. J., Boger, F., Feder, J., Jøssang, T. & Meakin, P. *Phys. Rev. A* (in the press).
21. Kapitulinik, A., Aharony, A., Deutscher, G. & Stauffer, D. *J. Phys. A* **16**, L269-L274 (1983).
22. Aharony, A. in *Multicritical Phenomena* (eds Pynn, R. & Skjeltorp, A.) 309-327 (Plenum, New York, 1984).
23. Aharony, A. in *Directions in Condensed Matter Physics* (eds Grinstein, G. & Mazenko, G.) 1-50 (World Scientific, Singapore, 1986).
24. Stauffer, D. *Introduction to Percolation Theory* (Taylor and Francis, London, 1985).
25. Nittmann, L. & Stanley, H. E. *Nature* **321**, 663-668 (1986).
26. Sherwood, J. D. & Nittmann, J. *J. Phys., Paris* **47**, 15-22 (1986).
27. DeGregoria, A. J. *Phys. Fluids* **28**, 2933-2935 (1985).
28. Murat, M. & Aharony, A. *Phys. Rev. Lett.* **57**, 1875-1878 (1986).
29. Nienhuis, B. *J. Phys. A* **15**, 199-213 (1982).
30. Herrmann, H. J., Hong, D. H. & Stanley, H. E. *J. Phys. A* **17**, L261-L266 (1985).
31. Laidlaw, D., MacKay, G. & Jan, N. *J. statist. Phys.* **46**, 507-515 (1987).
32. Bonnet, J. & Lenormand, R. *Rev. Inst. Fr. Pet.* **42**, 477-480 (1977).
33. Stanley, H. E. *J. Phys. A* **10**, L211-L220 (1977).
34. Coniglio, A. *Phys. Rev. Lett.* **46**, 250-253 (1981); *J. Phys. A* **15**, 3829-3844 (1982).
35. Gefen, Y., Aharony, A., Mandelbrot, B. B. & Kirkpatrick, S. *Phys. Rev. Lett.* **47**, 1771-1774 (1981).
36. Kirkpatrick, S. *AIP Conf. Proc.* **58**, 79-86 (1974).
37. Meakin, P., in *On Growth and Form* (eds Stanley, H. E. & Ostrowsky, N.) 111-135 (Martinus Nijhoff, Boston, 1986).
38. Eden, M. *Proc. 4th Berkeley Symp. on Math. Stat. and Prob.* **4**, 223-235 (1961).

LETTERS TO NATURE

Shock-compression fabrication of high-temperature superconductor/metal composite monoliths

L. E. Murr*†, A. W. Hare† & N. G. Eror*

* Department of Materials Science and Engineering, Oregon Graduate Center, Beaverton, Oregon 97006, USA
† Northwest Technical Industries, Sequim, Washington 98382, USA

In spite of the recent spectacular achievements in developing high-temperature (transition temperature $T_c > 77$ K) superconductivity in systems generically represented by $ABa_2Cu_3O_{7-x}$ (where $A = Y, La, Nd, Sm, Eu, Gd, Ho, Er, Lu$)^{1,2}, it has been difficult to prepare large powder sample batches (>0.5 kg). More importantly, efforts to consolidate or sinter superconducting powder samples or to place powders (such as Y-Ba-Cu-O) in useful environments or monolithic structures which are not brittle and friable, and the fabrication of high-temperature superconducting products have been frustrating at best³. Based on the process integration of several fundamental principles involving shock-compression science^{4,5}, we have successfully fabricated com-

posite/laminate monoliths containing shock-compacted, consolidated and bonded Y-Ba-Cu-O superconducting powder channels (with $T_c \geq 90$ K) in a solid copper matrix⁶. Using ammonium-nitrate-based explosives (with detonation velocities, V_D , adjusted to be $\sim 1,800$ m s⁻¹), we have progressed from shock-consolidated, superconducting $YBa_2Cu_3O_7$ powder-channel rings with mean diameters of 4.4 cm and 4 mm² cross-sections in explosively (implosive shock-wave) clad solid copper matrices, to monoliths from which copper matrix rings have been cut containing shock-consolidated, superconducting $YBa_2Cu_3O_7$ powder channels with mean diameters of 24 cm (with 2.5 cm width) and cross-sections of 1.85 cm².

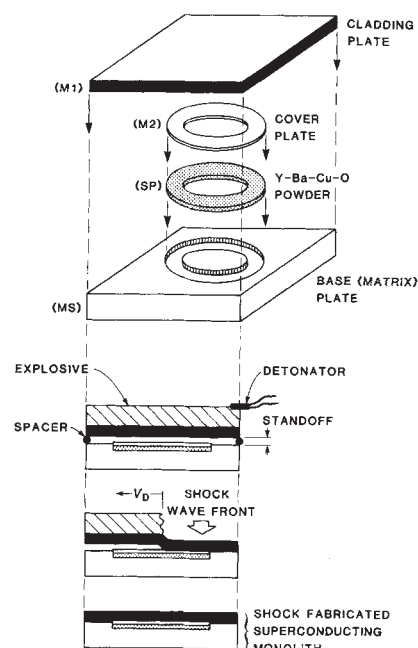
Figure 1 illustrates the development of these explosively (shock-wave) fabricated superconducting monoliths. By 'monolith' we mean a continuous (albeit composite), solid mass. In the context of Fig. 1 this means that the superconducting powder is dynamically compacted and consolidated by an explosively generated shock wave, and bonded to a metal (copper) matrix which consists of explosively bonded (or clad) planar laminates. There are at least four regimes in the simplest monolith prefabrications: the superconducting powder (SP) lightly compacted in a channel in a metal substrate (MS), a metal cover/secondary driver plate (M2), and a primary driver (cladding) plate (M1). At the conclusion of the shock event, all four regimes are bonded, and the powder is consolidated and bonded within the monolith.

Figure 2 shows typical results for the fabrication of small superconducting rings in copper-clad laminations. A weak Meissner effect was observed for shock-compression-fabricated monoliths such as that shown in Fig. 2 in liquid oxygen ($T = 90$ K), and a strong Meissner effect was observed in liquid nitrogen ($T = 77$ K). Rare-earth magnets can be levitated ~ 1 mm above the shock-consolidated, superconducting Y-Ba-Cu-O powder-channel region in the ring monolith shown in Fig. 2c and d, which had ~ 1 mm of solid copper covering the encapsulated superconducting channel structure in Fig. 2c, and 2 mm of solid copper covering the superconducting channel in Fig. 2d.

To produce large superconducting-channel cross-sections in copper-matrix monoliths, large powder batch production (>0.5 kg) was required. The $\text{YBa}_2\text{Cu}_3\text{O}_7$ superconducting powder was prepared from Y_2O_3 , BaCO_3 and CuO . The initial powders were milled in a vibratory mill to submicrometre size before heating. The milled powders were pressed to pellets of 5.95-cm diameter and heated in flowing oxygen at 870°C for up to 17 h. The pellets were reground in a mortar and pestle and heated in flowing oxygen at 950°C for 4 h. This process was repeated several times, with a final heating of the powder at 980°C , followed by an anneal in flowing oxygen at 300°C for 3 h and furnace cooling. The final powder was lightly ground ($\sim 2\text{--}200\ \mu\text{m}$).

The explosively fabricated, superconducting ring monoliths shown in Fig. 2 demonstrate a number of significant features, including the ability to incorporate and encapsulate large high-temperature superconducting regimes in 'engineering monoliths' or what we call manufacturing environments, in which a variety of metals and alloys can provide a supporting matrix and an encapsulation for the rather reactive, brittle and friable characteristics of the $\text{YBa}_2\text{Cu}_3\text{O}_x$ ($6 < x < 7$) powders and related

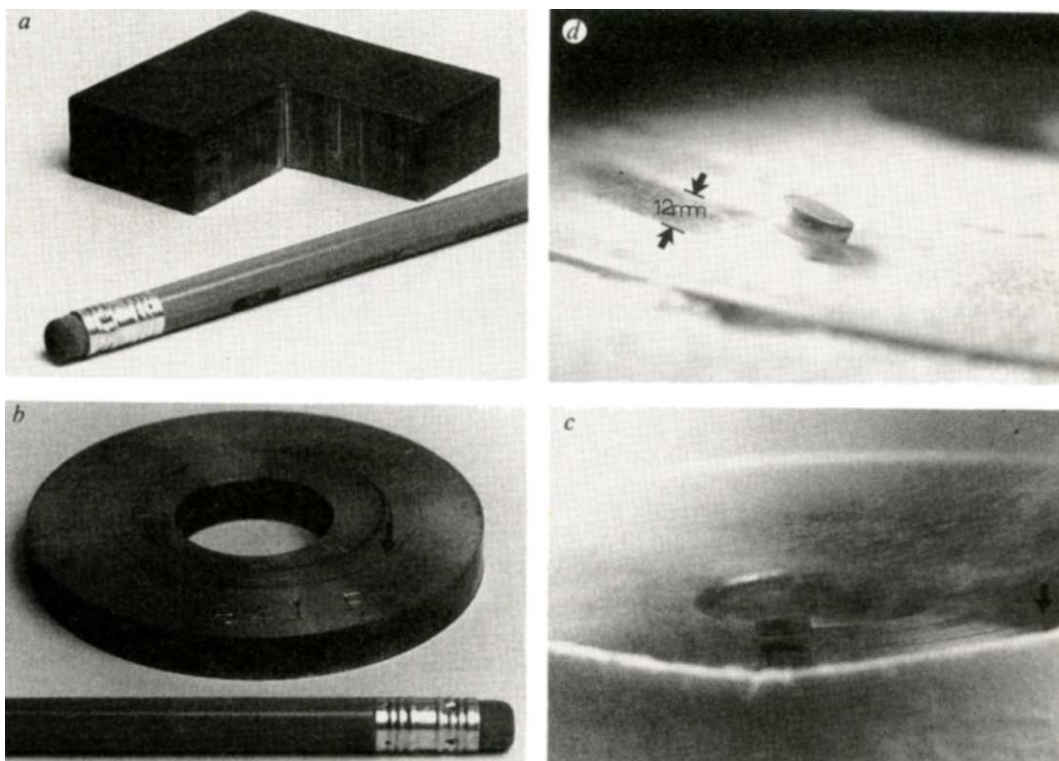
Fig. 1 Schematic view of the components and process configuration for the explosive (shock-wave) fabrication of a superconducting Y-Ba-Cu-O powder/metal (composite) monolith. In this simple schematic, a superconducting powder-channel ring is consolidated and encapsulated into a continuous, composite, explosively bonded metal monolith. The shock wave and cladding are effected by a flyer or cladding plate (M1). The cover plate (M2) is used to pre-compact the powder and acts as a secondary flyer or cladding plate which is welded to the primary flyer (M1) and bonded to the consolidated powder channel. The cladding plate M1 is also welded to the base plate. The four monolith regimes or domains can be of different materials, but here, copper was used throughout.



superconducting copper oxide powders. These 'superconducting engineering monoliths' can be machined, processed, welded and bonded, and subjected to a variety of metallurgical, manufacturing processes.

In addition, the shock-wave fabrication illustrated schemati-

Fig. 2 Examples of copper monoliths fabricated as in Fig. 1, containing consolidated Y-Ba-Cu-O superconducting powder-channel rings. a, Solid copper, superconductor-ring-containing fixture cut from larger monolith and centre-cut to expose the $1\text{ mm} \times 4\text{ mm}$ superconducting powder channel (black). The integrity of the monolith as depicted schematically in Fig. 1 is clearly evident. Note vertical band-saw marks. b, Early superconducting-channel-containing copper ring fixture cut and machined from an explosively fabricated monolith as shown in a. The machining was done off-centre and the channel edge was exposed slightly, creating a 'crack' which has developed over a month of cycling between liquid-nitrogen temperature and room temperature. c, Small rare-earth magnet levitating above the liquid-nitrogen-cooled ring shown in b. The arrows in b and c provide a reference point on the surface. The Y-Ba-Cu-O consolidated powder channel is slightly less than 1 mm below the copper metal surface at the channel centre. d, Large rare-earth magnet levitating over large, liquid-nitrogen-cooled, superconductor-containing, explosively fabricated copper ring monolith. An aluminized label sticking to the surface provides a dimensional reference (12 mm) as shown. The Y-Ba-Cu-O consolidated powder channel in this ring is 2.5 cm wide, ~ 0.7 cm thick, has an outer diameter of 25 cm, and an inner diameter of 20 cm. This channel contains ~ 0.55 kg of $\text{YBa}_2\text{Cu}_3\text{O}_7$ superconducting powder. The superconducting channel here lies below ~ 2 mm of copper.



cally in Fig. 1 does not involve significant heating ($T < 373$ K), as the peak shock pressures experienced by the powders during consolidation do not normally exceed ~ 10 GPa, and typical pressures achieved are estimated to be between 4 and 8 GPa for $V_D = 1,800$ m s $^{-1}$. Furthermore, we have used a distribution of fine grain sizes (2–200 μ m) to reduce the strain heating during dynamic compaction which occurs during void collapse, and which depends upon void size and distribution (ref. 7 and K. P. Staudhammer, L.E.M. and K. A. Johnson, manuscript in preparation).

We have shown in preliminary research that conventional approaches and other attempts to bond $\text{YBa}_2\text{Cu}_3\text{O}_7$ powders to copper tubes—to fabricate 'wire' forms, for example—are catastrophic failures when processing temperatures exceed ~ 750 °C for even short periods of time (≤ 0.5 h). This occurs because of the required, oxygen-rich environment, in which rapid oxygen diffusion along copper grain boundaries converts the copper to copper oxide. The resulting monoliths are extremely brittle, possess little strength or integrity, and are prone to intergranular fracture.

The explosively fabricated high-temperature superconducting monoliths illustrated in Figs 1 and 2 not only have the engineering advantages alluded to above, but also prospects for enhancement of the superconducting parameters as a result of shock-induced defects, particularly dislocations 4,8 . Although we have not specifically measured the dislocation density in the consolidated $\text{YBa}_2\text{Cu}_3\text{O}_7$ powder channels, typical changes in the dislocation density in commercial, solid copper subjected to mean peak pressures between ~ 4 and 8 GPa have been measured to be of the order of 10^3 cm $^{-2}$ (from 10^7 to 10^{10} cm $^{-2}$) 8 . Murr and Hiskey 9 have also measured a change in the dislocation density of shock-compressed, solid natural chalcocopyrite (CuFeS_2) of 10^2 cm $^{-2}$ (from 10^8 to 10^{10} cm $^{-2}$) at a mean peak shock pressure of 2 GPa, and 10^3 cm $^{-2}$ (from 10^8 to 10^{11} cm $^{-2}$) at 18 GPa. This provides evidence for plastic flow in ceramic-like materials similar to that observed in the high-pressure loading of copper and other metals alluded to above. Indeed, shock-wave-induced dislocations have been observed in classical ceramics such as alumina, including powdered materials 4,5 . Plastic deformation within the shock front will result in dislocations which could enhance mechanical bonding of the powder particles and could also promote enhancement of the superconducting parameters (particularly critical current and critical field). On the other hand, higher dislocation densities in exposed, consolidated powders could accelerate degradation because of enhanced reactivity, as demonstrated for shock-loaded chalcocopyrite in the work of Murr and Hiskey 9 . Plastic deformation of type II (low-temperature) superconducting alloys has been shown to favourably influence the behaviour of lower and upper critical field with temperature, as well as the critical current density, even when no discernible improvement in T_c was detected 10,11 . Hydrostatic pressure has been shown to have little effect on the superconducting state of the Y–Ba–Cu–O system up to peak pressures of 2 GPa (ref. 12), in contrast to the K_2NiF_4 -phase La–Ba–Cu–O and La–Sr–Cu–O systems where the T_c of La–Ba–Cu–O was enhanced at a rate of $>10^{-3}$ kbar $^{-1}$ (from ~ 37 K to 56 K) 1,13 .

We are currently shock-wave fabricating simple, straight bars or monolith composites of aluminium containing consolidated $\text{YBa}_2\text{Cu}_3\text{O}_7$ powder channels, to perform direct measurements of critical current density and other parameters. In preliminary measurements, we have induced persistent flux into rings like that shown in Fig. 2b, and have measured pulsed d.c. currents of 15 A through superconductor cross-sections of 4 mm 2 without any excursions from the superconducting state.

We have also cut into the superconducting channels around their circumference in order to demonstrate that the rings are continuous, and densely consolidated. These channels can be machined.

We thank M. Strasik, J. Javadpour, C. Niou and T. Monson

at the Oregon Graduate Center and D. Brasher at Northwest Technical Industries for technical assistance.

Received 6 August; accepted 17 August 1987.

1. Chu, C. W. *et al.* *Phys. Rev. Lett.* **58**, 405–407 (1987).
2. Wu, M. K. *et al.* *Phys. Rev. Lett.* **58**, 908–911 (1987).
3. Fletcher, F. J. *Metals* May (1987).
4. Meyers, M. A. & Murr, L. E. (eds) *Shock Wave and High-Strain-Rate Phenomena in Metals*, (Plenum, New York, 1981).
5. Murr, L. E., Meyers, L. E. & Staudhammer, K. P. (eds) *Metallurgical Applications of Shock-Wave and High-Strain-Rate Phenomena* (Dekker, New York, 1986).
6. Murr, L. E., & Hare, A. W. *US and International Patent Application* (29 July, 1987).
7. Murr, L. E., Shankar, S., Hare, A. W. & Staudhammer, K. P. *Scripta metall.* **17**, 1353–1359 (1983).
8. Murr, L. E. in *Shock Waves and High-Strain-Rate Phenomena in Metals* (eds Meyers, M. A. & Murr, L. E.) 607–673 (Plenum, New York, 1981).
9. Murr, L. E. & Hiskey, J. B. *Metall. Trans.* **12B**, 255–267 (1981).
10. Livingstone, J. D. & Schadler, H. W. *Progress in Materials Science* Vol. 12 (ed. Chalmers, B.) 185–287 (1964).
11. Geballe, T. H. in *Critical Materials Problems in Energy Production* (ed. Stein, C.) (Academic, New York, 1976).
12. Hor, P. H. *et al.* *Phys. Rev. Lett.* **58**, 911–912 (1987).
13. Chu, C. W., Hor, P. H., Meng, R. L., Gao, L. & Huang, Z. J. *Science* **235**, 567–568 (1987).

Tetrakis(methyltelluro)tetrathiafulvalene (TTeC $_1$ -TTF), a high-mobility organic semiconductor

H. Inokuchi*, K. Imaeda*, T. Enoki*, T. Mori*,
Y. Maruyama*, G. Saito†, N. Okada†, H. Yamochi†,
K. Seki‡, Y. Higuchi§ & N. Yasuoka§

* Institute for Molecular Science, Okazaki 444, Japan

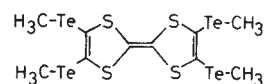
† The Institute for Solid State Physics, University of Tokyo, Tokyo 106, Japan

‡ Department of Materials Science, Faculty of Science, Hiroshima University, Hiroshima 730, Japan

§ Basic Research Laboratory, Himeji Institute of Technology, Himeji 671-22, Japan

We have found a class of highly conductive single-component organic semiconductors; the tetrakis(alkylthio)tetrathiafulvalenes, and have given these systems the descriptive title of 'molecular fastener' 1,2 . In a programme to synthesize single-component molecular assemblies with high conductivity, we then discovered a new organic semiconductor, tetrakis(alkyltelluro)tetrathiafulvalene, of electrical resistivity 8.1×10^4 Ω cm, having high charge mobility, 20–30 cm 2 V $^{-1}$ s $^{-1}$. We explain this unusually low resistivity in terms of molecular packing and stacking manner within the crystal. The central skeleton of this compound, tetratelluro-tetrathiafulvalene moiety ($\text{C}_6\text{Te}_4\text{S}_4$), is almost planar and is regularly stacked in the form of column. Along the stacking axis, tellurium atoms in neighbouring columns come close each other and form zigzag chains. The distance between those two Te atoms is 3.644(2) Å, which is significantly shorter than van der Waals distance (4.12 Å). It is expected that zigzag chalcogen chains formed from quasi-covalent bonds.

The molecular structure of this class of compounds, tetrakis(methyltelluro)tetrathiafulvalene (abbreviated to TTeC $_1$ -TTF), is shown below. A series of tetrakis(alkyltelluro)tetrathiafulvalenes were synthesized using the method developed by



Aharon-Shalom *et al.* 3,4 . TTF was added to the tetrahydrofuran solution of lithium diisopropylamide at -78 °C, stirred and then amorphous tellurium powder was added. The temperature of the mixture was gradually raised to 0 °C. After the dissolution

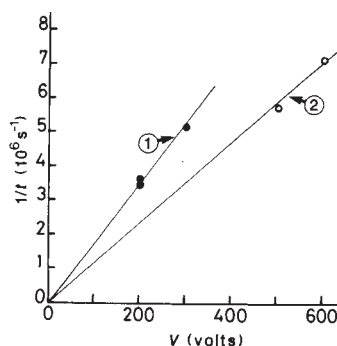


Fig. 1 The electric-field dependence of the reciprocals of transit-time. The carrier mobility can be derived from $\mu = d^2/t.V$; where d is the crystal thickness, V is the applied voltage and t is the transit-time. Curve (1), $\mu_h = 28.5 \pm 0.5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$; curve (2), $\mu_e = 18.6 \pm 0.4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.

of the Te powder, methyl iodide was added to the recooled solution (-78°C), which was again stirred. The reaction mixture was then warmed gradually to room temperature, after which water was added. The precipitated crude product was washed with water and methanol and recrystallized from CHCl_3 or petroleum ether to obtain brownish-yellow transparent crystals with a melting point of 175°C . This yield of the process is about 66%. Samples of high purity were subsequently obtained by careful recrystallization, having the appearance of dark orange needles.

The direct current (d.c.) conductivities of the needle-like single crystals ($0.3 \times 0.3 \times 2 \text{ mm}$) were measured by a four-probe method as a function of temperature. Four $5 \mu\text{m}$ -diameter gold wires were attached to the crystal with gold paste and the conductivity measurement was carried out in a vacuum vessel at 10^{-4} Pa . The dark conductivity of $\text{TTeC}_1\text{-TTF}$ at room temperature was $1.2 \times 10^{-5} \text{ Scm}^{-1}$ in b -direction, with an activation energy $E_a = 0.25 \text{ eV}$. This conductivity is extraordinarily high among single-component organic semiconductors.

We also established the charge mobility of this organic semiconductor. The drift mobility was measured using surface-type time-of-flight technique⁵. A d.c. voltage within the range from 100 V to 600 V was applied across the specimen, having 0.4 mm length. The light source for the carrier photo-generation was a N_2 -gas laser (wavelength, 337.1 nm; pulse-width, $\leq 1 \text{ ns}$). The photo-response was detected using a storage oscilloscope (Tektronics model 7834). Figure 1 shows the relation between the reciprocal of the time of flight and the applied voltage. From this relation, we were able to measure the charge mobility; $\mu_h = 28.5 \pm 0.5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $\mu_e = 18.6 \pm 0.4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. This μ -value is very large compared with those of ordinary organic semiconductors ($\sim 1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ for anthracene, for example).

The molecular packing and stacking manner of $\text{TTeC}_1\text{-TTF}$ is illustrated in Fig. 2. This structure was confirmed by X-ray diffraction measurements where intensity data were collected on an automated single crystal diffractometer using monochromatized MoK_α radiation. The positions of the tellurium atoms were identified using a MULTAN 80 program package and those of other atoms were found by a weighted Fourier technique. The atomic positions were finally established using a full matrix least-squares refinement procedure. The crystallographic R -value was reduced to 0.050, based on 1,199 non-zero reflections. Crystal data and other parameters related to the structure analysis are as follows: Space group $\text{P2}_1/\text{n}$, $a = 13.769(3) \text{ \AA}$, $b = 5.480(3) \text{ \AA}$, $c = 12.245(2) \text{ \AA}$, $\beta = 90.49(1)^\circ$, $V = 923.8(5) \text{ \AA}^3$ and $Z = 2$.

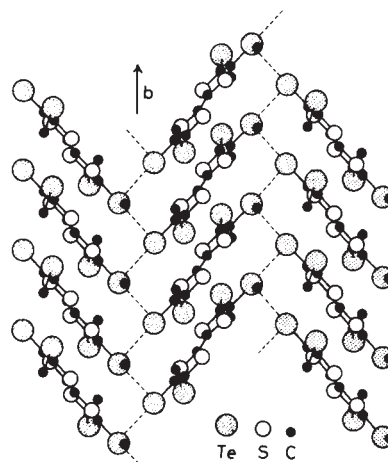


Fig. 2 The packing of $\text{TTeC}_1\text{-TTF}$ molecules in a crystal.

As mentioned above, along the stacking axis, tellurium atoms in neighbouring columns come close each other and form zigzag chains. Therefore the zigzag chains $-\text{Te}-\text{Te}-\text{Te}-\text{Te}-$, shown as dotted lines in Fig. 2, should have a quasi-covalent character similar to that observed in elemental Te. The short Te-Te distance in crystalline Te ($2.84 \text{ \AA}$) is associated with strong covalent interactions and high electrical conductivity, ($1 \Omega\text{cm}$). These facts lead us to the conclusion that the high conductivity and the large electron mobility of $\text{TTeC}_1\text{-TTF}$ occurs via electron transport through the zigzag Te-Te-Te quasi-covalent chains in the crystal predominantly. The detailed structure will be reported elsewhere.

This conclusion is reinforced by the low conductivity of tetrakis(ethyltelluro)-tetrathiafulvalene ($\text{TTeC}_2\text{-TTF}$) ($\rho = 2.3 \times 10^9 \Omega\text{cm}$) at room temperature, where the Te-Te distance between neighbouring columns is fairly large, $3.987(1) \text{ \AA}$ ($a = 14.017(4) \text{ \AA}$, $b = 9.021(2) \text{ \AA}$, $c = 9.239(5) \text{ \AA}$, $\beta = 104.34(3)^\circ$ and $V = 1131.8(1) \text{ \AA}^3$).

Although there is a report on a similarly close Te-Te contact in hexamethylenetetratellurafulvalene, where the molecules form columns stacked alternately with Te-Te distance of 3.583 and $3.743 \text{ \AA}$ (ref. 6), the conductivity of this compound was very low, due to dimerization effects.

The first series of reports on organic semiconductors which appeared around 1950 were on single-component systems such as polycyclic aromatics⁷, phthalocyanine⁸ and some dyes⁹. Since 1954 multicompound systems such as charge transfer complexes and doped polymers have been developed¹⁰ and this group now includes organic conductors and even organic superconductors¹¹. However, semiconducting single-component organic compounds are likely to be much more suitable for use as molecular devices because they can easily be made into thin film.

The introduction of covalency to van der Waals forces can be used as a general strategy to produce highly conductive single-component organic semiconductors.

Received 6 April; accepted 15 July 1987.

1. Inokuchi, H. *et al. Chem. Lett.* 1263-1266 (1986).
2. Seki, K. *et al. JCS Faraday Trans. 2*, 82, 1067-75 (1986).
3. Ahron-Shalom, E., Becker, J. Y., Bernstein, J., Bittner, S. & Shaik, S. *Tetrahedron Lett.* 26, 2783-2786 (1985).
4. Okada, N., Yamochi, H., Shinozaki, F., Oshima, K. & Saito, G. *Chem. Lett.* 1861-1864 (1986).
5. Maruyama, Y. & Yamada, T. *J. non-Cryst. Solids* 28, 143-145 (1978).
6. Carroll, P. J. *et al. JCS Chem. Commun.* 1316-1318 (1982).
7. Akamatsu, H. & Inokuchi, H. *J. chem. Phys.* 18, 810-811 (1950).
8. Eley, D. D. *Nature* 162, 819 (1948).
9. Vartanyan, A. T. *Zhru. Fhim. Khim.* 24, 1361-1370 (1950).
10. Akamatsu, H., Inokuchi, H. & Matsunaga, Y. *Nature* 173, 168-169 (1954).
11. Jerome, D., Mazaud, A., Ribault, M. & Bechgaard, K. *J. Phys. Lett.* 41, 95-98 (1980).

Aftershocks of deep earthquakes do not occur preferentially on nodal planes of focal mechanisms

Cliff Frohlich & Raymond J. Willemann*

Institute for Geophysics, University of Texas at Austin,
Austin, Texas 78713, USA

* Los Alamos National Laboratory, Los Alamos,
New Mexico 87545, USA

Some earthquakes with focal depths exceeding 70 km possess one or more aftershocks which are well recorded by stations at teleseismic distances. Precise determination of relative vectors separating the aftershocks and ~60 initial events finds the vector directions are distributed uniformly with respect to the nodal plane of main shock focal mechanisms. When focal mechanisms are available for both an initial event and an aftershock, the mechanisms are occasionally significantly different from one another. These results suggest that the failure process of deep earthquakes is not slip occurring along a simple, planar surface. Our results contradict several previous analyses of aftershocks because other investigators did not recognize that half of the focal sphere lies within 16° of one of the nodal planes.

Although about a fifth of all the world's teleseismically reported earthquakes have focal depths exceeding 70 km, the nature of the failure process that allows these events to occur remains one of the outstanding problems in rock mechanics. Several studies have determined precise relative locations for groups of deep events occurring over periods of years and find that deep earthquakes often appear to form elongate or planar groups lying approximately parallel to the nodal planes of focal mechanisms of nearby events¹⁻³. These results have been interpreted to suggest that deep earthquakes must occur as failure along a planar surface. A different test of this hypothesis is to study the relationship between individual events and their aftershocks, which are an integral part of the failure process.

Unlike large shallow earthquakes, deep earthquakes seldom possess well-developed aftershock sequences having numerous events⁴⁻⁶. However, some of these deep-focus events occur as 'doublets' or 'multiplets', events which occur too close to one another in space and time to be causally unrelated^{7,8}. Recently, Frohlich⁸ found that ~10% of all teleseismically reported deep events having magnitudes of 5.0 or larger possessed at least one true aftershock, a distinct event statistically demonstrable as related to the initial event, and occurring a minute or more after the initial event. Furthermore, detailed analysis of the seismic waveforms of large, deep earthquakes reveals that the sources of most are complex, consisting of two or more rupture sub-events, occurring within a minute or less, and within tens of km of one another^{7,9-19}. For deep earthquakes, analysis of the locations of rupture subevents and true aftershocks provides the best available information concerning the spatial and temporal scale of the failure process which allows these events to occur within the mantle.

The primary objective of this report is to investigate the relationship between the observed focal mechanisms of deep events and the orientations of separation vectors linking the initial events to true aftershocks and rupture subevents. With one exception²⁰, previous studies of this phenomenon have looked at only a few events. Also, comparison of their results⁷⁻²⁰ is difficult because each used slightly different methods for relocating events, and for determining the uncertainties in their locations. To obtain a relatively large data set of uniform quality suitable for statistical analysis, we used relative locations reported by other investigators for 5 sequences, and we relocated 20 sequences investigated by others using the published relative arrival times. We augmented these data with relocations of 26

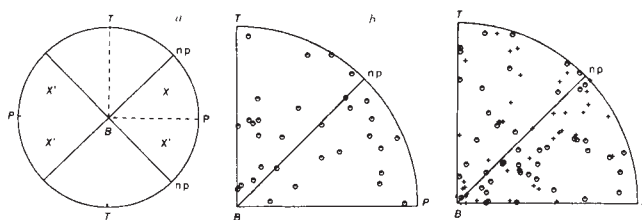


Fig. 1 Comparison of aftershock directions and orientation of focal mechanisms. A focal mechanism uses energy from the earthquake to determine three perpendicular axes, the axis of maximum compressional stress at the earthquake source (P axis), the axis of maximum tensional stress (T axis) and the null axis (B axis). Note that the directions appear to be distributed uniformly, with no tendency to cluster along the nodal plane (np). *a*, All the directions X' and X are equidistant from the P , T and B axes and the nearest nodal plane. Thus, when the nodal planes are indistinguishable a quarter hemisphere (within dashed lines) is sufficient to show the possible relationships of directions to the focal mechanism. *b*, Observed directions from initial event to rupture subevents compared to orientation of focal mechanism of the initial event. *c*, Observed directions from initial events to true aftershocks compared to orientation of focal mechanism of the initial event. Open circles are directions evaluated in our study, pluses are those reported by Oike²⁰.

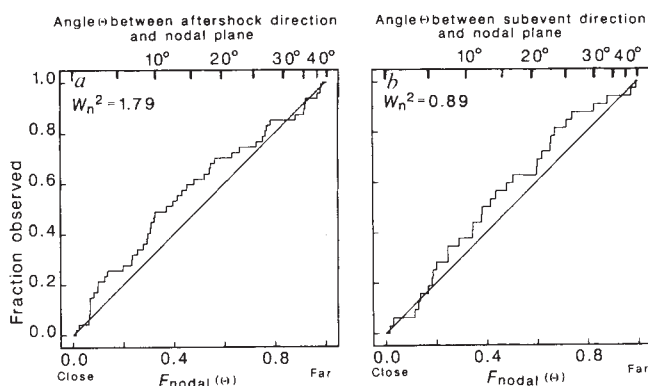


Fig. 2 For data in Fig. 1 (open circles only) cumulative distribution of angles Θ_k separating nodal planes and aftershock directions (step function), compared to distribution expected for uniformly distributed directions (diagonal straight line). For each of n directions Θ_k , the y value of the step is plotted between $(k-1)/n$ and k/n ; while the function $F_{\text{nodal}}(\Theta_k)$ on the x axis, is chosen to map uniformly distributed directions into a uniform distribution. Because the Anderson-Darling statistic W_n^2 has a value close to 1.0, the distributions of angles for (a) rupture subevents, and (b) true aftershocks are not significantly different from uniform²¹.

sequences, using arrival times read by us, and relocations of 11 sequences with times reported in the *Bulletin of the International Seismological Centre*. To display the orientations of relative location vectors with respect to focal mechanisms, we plotted them on a quarter-hemisphere, equal-area projection (Fig. 1).

The results indicate that neither rupture subevents nor true aftershocks occur preferentially along directions lying parallel to the nodal planes. A visual inspection of an equal-area projection of the data (Fig. 1) suggests the data are distributed uniformly over the focal sphere, in accord with a statistical analysis using Bingham statistics²¹. Furthermore, if we directly compare the observed distribution of angles between the direction vectors and the nearest nodal plane to the distribution expected for random directions on a focal sphere (Fig. 2) we find the observed distribution is not significantly different from uniform, as measured by Anderson-Darling statistics²¹.

Because our results concerning the relative locations of aftershocks apparently contradict many previous investigations⁹⁻²⁰, we now consider how our analysis differs from that of the previous investigators. First, none of the previous investigators performed a statistical analysis of the directional data. They also did not state explicitly under what conditions a vector

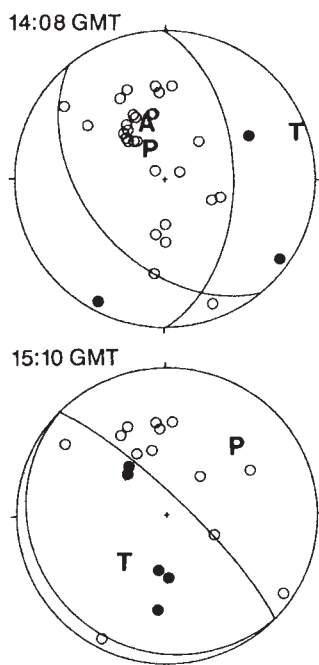


Fig. 3 Focal mechanisms of deep earthquakes are occasionally very different from the mechanisms of their aftershocks. Top, lower hemisphere, equal-area projection for the event of 25 October 1973 at 14:08 GMT, with a magnitude m_b of 6.1, occurring at a depth of 517 km beneath Peru-Bolivia. The letter A indicates direction of the aftershock. Bottom, mechanism for the aftershock of 15:10 GMT, with a magnitude m_b of 5.2, occurring 36 km distant from the initial shock. Open circles are dilatations, and closed circles are compressions, read by the authors from microfiche copies of seismograms at stations around the world. The mechanisms are determined from this data using the objective search program written by Reasenberg & Oppenheimer³⁰.

direction lay 'close to' a nodal plane, and apparently did not recognize that 50% of all directions on a focal sphere lie within 15.8° of at least one of the nodal planes. Because they interpreted events individually, and because the determination of any particular focal mechanism and aftershock direction may each be uncertain by perhaps 10° , they were seldom able to rule out the possibility that the direction lay along the nodal plane. Oike's²⁰ study of many aftershock sequences differed from ours in that he did not use statistical criteria either to identify aftershocks or to evaluate clustering of the computed directions. When we analyse his published results (Fig. 1) using our statistical methods, we find only weak evidence for clustering along nodal planes. In particular, most of the apparent clustering is due to ten aftershocks of a single large and anomalous event which occurred at 132-km depth on 1 December 1966.

The failure process of deep earthquakes must differ from that of shallow earthquakes, because laboratory studies show that most materials will undergo ductile, rather than brittle failure at mantle temperatures and pressures²²⁻²⁴. Successful theories of shallow brittle failure depend on the opening and propagation of small cracks²⁵, which cannot exist at mantle pressures. But suggestions that deep earthquakes might be simple implosions caused by rapidly running phase transitions are not consistent with the observed predominance of the double-couple component of the source, and the near absence of isotropic radiation^{26,27}. Thus, most previous observers^{1,5,6} have assumed that failure, as represented by rupture subevents and true aftershocks, occurred primarily as slip along a simple, planar surface.

We find that individual rupture subevents and true aftershocks may not occur preferentially along nodal planes. This result suggests that the failure occurs in a non-planar zone distributed around the location of the initial event. For example, the failure zone may occur at the boundaries of a zone undergoing a phase

transition, and stressing neighbouring regions. Alternatively, the failure may occur along curved or contorted surfaces, which are non-planar because either the regional stresses or the material properties in the source region are inhomogeneous over the spatial scale of the failure process. The present result is consistent with several investigations^{28,29} which report that at least occasionally focal mechanisms of deep earthquakes differ markedly from the mechanisms of true aftershocks (Fig. 3).

Received 30 April; accepted 22 June 1987.

1. Billington, S. & Isacks, B. *Geophys. Res. Lett.* **2**, 63-66 (1975).
2. Billington, S., Isacks, B. & Barazangi, M. *Geology* **5**, 699-704 (1977).
3. Giardini, D. & Woodhouse, J. H. *Nature* **307**, 505-509 (1984).
4. Page, R. J. *geophys. Res.* **73**, 3897-3903 (1968).
5. Kagan, Y. Y. & Knopoff, L. *Bull. seism. Soc. Am.* **70**, 1811-1822 (1980).
6. Prozorov, A. G. & Dziewonski, A. M. *J. geophys. Res.* **87**, 2829-2839 (1982).
7. Isacks, B. L., Sykes, L. R. & Oliver, J. *Bull. seism. Soc. Am.* **57**, 935-958 (1967).
8. Frohlich, C. J. *geophys. Res.* **92**, (in the press).
9. Oike, K. *Bull. Dis. Prev. Res. Inst.* **19**, 55-66 (1969).
10. Oike, K. *Bull. Dis. Prev. Res. Inst.* **20**, 145-182 (1971).
11. Choy, G. L. & Boatwright, J. *Bull. seism. Soc. Am.* **71**, 691-711 (1981).
12. Fukao, Y. *Phys. Earth planet. Inter.* **5**, 61-76 (1972).
13. Chung, W. & Kanamori, H. *Phys. Earth planet. Inter.* **13**, 85-96 (1976).
14. Furumoto, M. *Phys. Earth planet. Inter.* **15**, 1-12 (1977).
15. Koyama, J. *Phys. Earth planet. Inter.* **16**, 307-317 (1978).
16. Sasatani, T. *J. Fac. Sci. Hokkaido Univ. Geophys.* **6**, 301-384 (1980).
17. Rakers, E. & Muller, G. *J. Geophys.* **50**, 143 (1982).
18. Brustle, W. & Muller, G. *Bull. seism. Soc. Am.* **77**, 47-68 (1987).
19. Koyama, J. *Sci. Rep. Tohoku Univ. Geophys.* **23**, 83-101 (1975).
20. Oike, K. *Bull. Dis. Prev. Res. Inst.* **21**, 153-178 (1971).
21. Cox, D. R. & Lewis, P. A. W. *The Statistical Analysis of a Series of Events* 150 (Chapman & Hall, London, 1966).
22. Heard, H. C. in *Rock Deformation, Geol. Soc. Am. Mem.* **79**, (eds Griggs, D. & Handin J.) 193-226 (Waverly, Baltimore, 1960).
23. Orowan, E. in *Rock Deformation, Geol. Soc. Am. Mem.* **79**, (eds Griggs, D. & Handin, J.) 323-345 (Waverly, Baltimore, 1960).
24. Griggs, D. T. & Baker, D. W. in *Properties of Matter Under Unusual Conditions* in (ed. Mark, H.) 23-42 (Interscience, New York, 1969).
25. Dmowska, R. & Rice, J. R. in *Continuum Theories and Solid Earth Physics* (ed. Teisseyre, R.) (Elsevier, New York, 1983).
26. Silver, P. G. & Jordan, T. H. *Geophys. J.* **70**, 755-787 (1982).
27. Hinzen, K. *J. Geophys.* **59**, 112-120 (1986).
28. Isacks, B. & Molnar, P. *Rev. Geophys. Space Phys.* **9**, 103-174 (1971).
29. Schneider, J. F. thesis, Univ. Wisconsin, Madison (1984).
30. Reasenberg, P. & Oppenheimer, D. *US Geol. Surv. Open-File Rep.* No. 85-739 (US Geological Survey, Menlo Park, 1986).

Mechanisms for northwards dispersal of Sellafield waste

A. B. Mackenzie*, R. D. Scott* & T. M. Williams*†

* Scottish Universities Research and Reactor Centre, East Kilbride, Glasgow G75 0QU, UK

† Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow G1 1XL, UK

Almost all of the plutonium and americium discharged as liquid effluent from Sellafield is retained in the sediments of the north east Irish Sea¹⁻⁵, whereas the waste radiocaesium shows relatively conservative behaviour in seawater and is mostly dispersed into more distant waters with only a small fraction being retained in Irish Sea sediment⁶⁻¹². Sellafield discharges^{13,14} considered in the light of radionuclide geochemistry and radioactive decay and growth¹⁵ imply that, by the end of 1984, the sediments of the north-east Irish Sea contained about 6×10^2 TBq each of ^{239,240}Pu and ²⁴¹Am and of the order of 10^3 TBq of ¹³⁷Cs. This contamination, of globally significant magnitude^{16,17}, is largely concentrated in restricted areas of fine sediment off the Cumbrian coast close to Sellafield^{3,4,18} so it is important to identify processes which could lead to the dispersal of these nuclides to areas of potential human contact (such as intertidal sediments) over the long time scales appropriate to the half lives of ²⁴¹Am (458 yr), ²³⁹Pu (2.44×10^4 yr) and ²⁴⁰Pu (6.58×10^3 yr). Because no significant return of these actinides to solution in seawater is likely^{19,20}, most concern must be centred upon physical dispersal processes affecting the contaminated sediment. These are ill-defined, although Hunt²¹ contends that actinides incorporated in intertidal sediments near Sellafield

are subject to mixing with previously discharged waste and to dispersal, with removal from the surface sediments within one or two decades. Here we present the results of a study of radionuclide concentrations in surface (<0.5 cm depth) intertidal sediments collected during 1984–85 which provide the basis for an estimate of the extent of northwards dispersal of the contaminated sediment and an assessment of the importance of this process relative to transport of radionuclides in solution.

The samples were analysed by established methods^{22–24} and the results, presented in Table 1, show obvious discontinuities, both in the concentrations of the individual radionuclides and in the activity ratios, which suggest that the samples can (with the few exceptions discussed below) be grouped into three distinct classes corresponding to the geographical areas A, B and C delineated in Fig. 1. The broad features of this classification are summarized in Table 2 and we now attempt to relate them to the characteristics of the known source terms, namely Sellafield liquid effluent and fallout from atmospheric testing of nuclear weapons²⁵.

We assume that radionuclides are transported to the intertidal sediments studied by processes which are initially controlled by partitioning in the seawater–sediment system and subsequently by movement of contaminated water or sediment, so that the Sellafield discharge can be regarded as providing two separate source terms, namely particle-associated and soluble, with characteristic nuclide activity ratios resulting from the different chemical behaviour of the various nuclides. Hunt's results¹⁸ for silts from Newbiggin, near Sellafield, in 1984 have been taken to typify the particle-associated component and Hetherington's work¹, relating the $^{137}\text{Cs}/^{239,240}\text{Pu}$ activity ratio in filtered Irish Sea water to that of the Sellafield discharge, has been employed in conjunction with appropriate distribution coefficients (K_D)^{1,4,3,26} and the 1984 discharge figures to estimate the effects

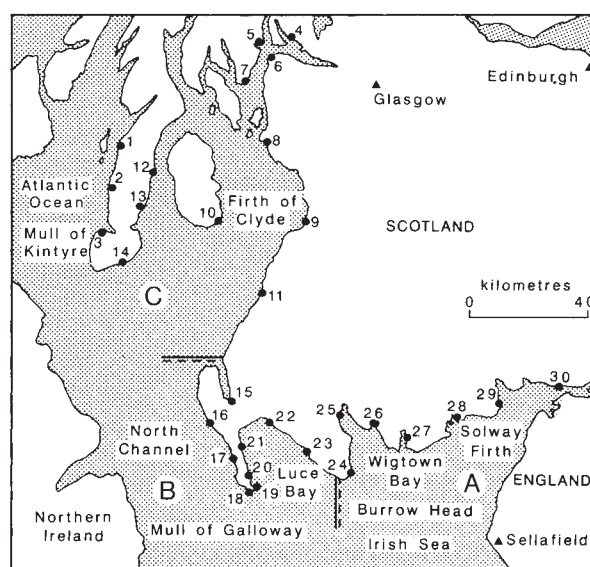


Fig. 1 Map of the area and locations of the sampling sites.

of the soluble component. The fallout figures are derived from those appropriate to soils in south-west Scotland^{17,27,28}.

Turning now to the analytical results, the grouping in Fig. 1 does not correspond to any simple variation of sedimentary facies between the three regions although, as would be expected within any one of the three, radionuclide concentrations are highest in the finest grained sediment. Sediments from region A have the highest concentrations in conjunction with isotopic

Table 1 Analytical results for surface intertidal sediments

Sampling site (date)	^{137}Cs	^{241}Am	$^{239,240}\text{Pu}$	^{238}Pu	Dominant sediment type
				$^{239,240}\text{Pu}$	
1 (4/85)	51 ± 3	1.3 ± 0.5	2.7 ± 0.2	0.19 ± 0.04	Sand, shell fragments
2 (4/85)	19 ± 3	0.7 ± 0.3	1.15 ± 0.07	0.19 ± 0.02	Sand, shell fragments
3 (4/85)	51 ± 3	<1.2	1.22 ± 0.08	0.18 ± 0.04	Sand, shell fragments
4 (9/84)	94 ± 3	<0.7	1.17 ± 0.08	0.16 ± 0.04	Sand, silt
5 (9/84)	108 ± 3	<0.5	1.43 ± 0.14	0.15 ± 0.02	Sand, silt
6 (9/84)	66 ± 2	<0.5	0.67 ± 0.05	0.15 ± 0.03	Sand, gravel
7 (9/84)	54 ± 2	<1.2	1.33 ± 0.10	0.18 ± 0.02	Sand
8 (9/84)	49 ± 2	<0.5	0.67 ± 0.05	0.18 ± 0.02	Sand, shell fragments
9 (9/84)	49 ± 2	<1.0	0.83 ± 0.07	0.16 ± 0.03	Sand
10 (9/84)	76 ± 2	<1.2	0.95 ± 0.08	0.16 ± 0.03	Sand, shell fragments
11 (9/84)	61 ± 3	<0.6	1.48 ± 0.12	0.19 ± 0.02	Sand
12 (4/85)	27 ± 3	<0.4	0.51 ± 0.04	0.18 ± 0.02	Sand, shell fragments
13 (4/85)	51 ± 3	<0.4	0.44 ± 0.03	0.14 ± 0.07	Sand
14 (4/85)	37 ± 1	0.7 ± 0.35	0.79 ± 0.06	0.16 ± 0.05	Sand
15 (4/85)	44 ± 3	0.8 ± 0.2	1.05 ± 0.07	0.18 ± 0.04	Sand
16 (4/85)	92 ± 2	6.2 ± 0.7	4.9 ± 0.3	0.18 ± 0.02	Sand, silt
17 (4/85)	65 ± 3	2.3 ± 0.3	2.5 ± 0.2	0.20 ± 0.02	Sand
18 (4/85)	97 ± 3	10 ± 1	5.4 ± 0.4	0.21 ± 0.01	Gravel, silt
19 (4/85)	58 ± 2	3.3 ± 0.7	2.7 ± 0.2	0.18 ± 0.02	Sand
20 (4/85)	75 ± 3	1.7 ± 0.6	2.1 ± 0.2	0.18 ± 0.02	Sand
21 (4/85)	62 ± 3	5.7 ± 0.9	4.3 ± 0.3	0.21 ± 0.03	Sand
22 (4/85)	98 ± 3	5.5 ± 1.0	5.4 ± 0.3	0.19 ± 0.02	Sand, shell fragments
23 (5/85)	99 ± 3	4.8 ± 1.6	4.3 ± 0.3	0.21 ± 0.02	Sand, silt, shell fragments
24 (5/85)	320 ± 6	60 ± 6	39 ± 3	0.21 ± 0.02	Sand, silt, shell fragments
25 (5/85)	1,030 ± 20	194 ± 12	132 ± 9	0.24 ± 0.02	Silt, sand
26 (5/85)	412 ± 8	91 ± 8	53 ± 4	0.21 ± 0.01	Sand, silt, shell fragments
27 (5/85)	1,170 ± 20	186 ± 12	124 ± 8	0.22 ± 0.01	Sand, silt, shell fragments
28 (5/85)	768 ± 15	137 ± 14	105 ± 7	0.25 ± 0.02	Silt, sand
29 (5/85)	431 ± 8	37 ± 7	31 ± 2	0.23 ± 0.02	Sand, silt, shell fragments
30 (5/85)	328 ± 6	15 ± 3	5.4 ± 0.4	0.29 ± 0.06	Sand, silt, shell fragments

The locations of the sampling sites are shown in Fig. 1. Radionuclide concentrations are in units of Bq kg^{-1} (dry). ^{134}Cs was also detected at low concentrations in all of the samples but the large uncertainties resulting from poor counting statistics preclude the use of ^{134}Cs measurements for quantitative interpretation.

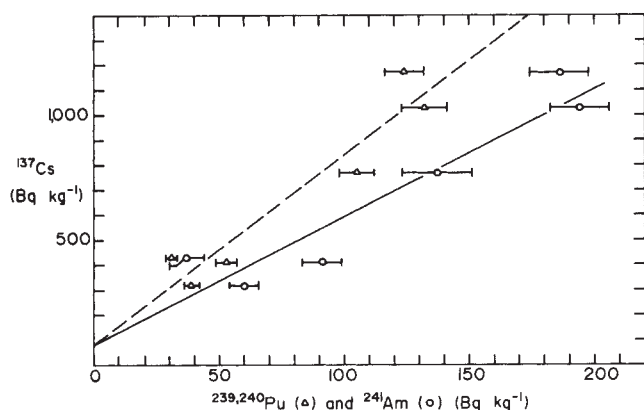


Fig. 2 The linear correlations between the ^{137}Cs activity and the $^{239,240}\text{Pu}$ and ^{241}Am activities in region A.

and nuclide activity ratios similar to those of the particle-associated component of the Sellafield waste but incompatible with those of the other two sources (with the exception of sample 30 which is probably influenced by local discharges from Chapelcross and is consistent with results reported by the Ministry for Agriculture, Fisheries and Food (MAFF)¹⁸ for the Chapelcross area). We therefore suggest that the transport of contaminated particulate material is the main mechanism of supply to region A and attribute the systematic differences which do exist between these sediments and those close to Sellafield (that is, the lower concentrations and higher $^{137}\text{Cs}/^{239,240}\text{Pu}$ activity ratios in region A) to dilution and mixing during transit coupled with the emerging influence of the soluble component as the distance from Sellafield increases.

An order of magnitude step change in radionuclide concentrations occurs at Burrow Head and the $^{241}\text{Am}/^{239,240}\text{Pu}$ values for area B again favour sediment transport, but the lower concentrations indicate that the process is much less efficient than in region A. The slight decrease in $^{238}\text{Pu}/^{239,240}\text{Pu}$ activity ratios, compared with those of region A, while supporting this interpretation, also suggests a small influence of atmospheric fallout or mixing with much older Sellafield waste. The $^{137}\text{Cs}/^{239,240}\text{Pu}$ and $^{137}\text{Cs}/^{241}\text{Am}$ activity ratios in region B are markedly higher

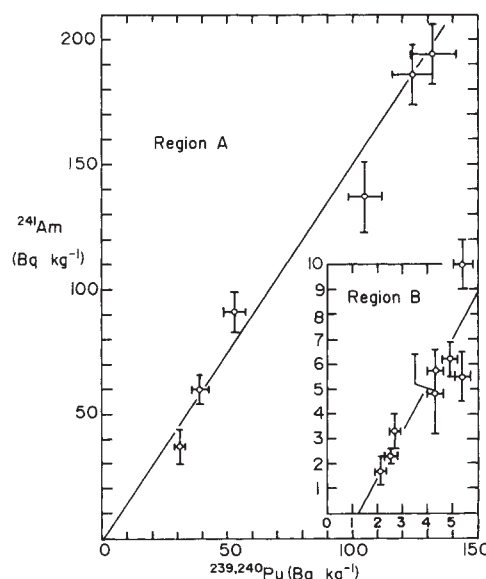


Fig. 3 The linear correlation between the $^{239,240}\text{Pu}$ and ^{241}Am activities in regions A and B.

than those in region A, which implies that soluble waste from Sellafield and possibly atmospheric fallout are contributing to the radiocaesium in region B. The actinide activity ratio, on the other hand, shows only a small decrease on passing from A to B and this is consistent with a single source of supply for the highly insoluble americium while allowing the possibility of a small contribution to plutonium from the soluble waste.

A further obvious change occurs between regions B and C: region C has the lowest observed concentrations and the decrease is most pronounced for the actinides. The low $^{241}\text{Am}/^{239,240}\text{Pu}$ in conjunction with high $^{137}\text{Cs}/^{239,240}\text{Pu}$ and $^{137}\text{Cs}/^{241}\text{Am}$ activity ratios indicates that transfer of particle-associated waste is not important in this area, while the average value of 0.17 for the $^{238}\text{Pu}/^{239,240}\text{Pu}$ activity ratio and the presence of low concentrations of ^{134}Cs imply that the soluble component of the Sellafield waste plus atmospheric fallout are the main sources of radionuclide supply to the intertidal sediments of region C. Samples 1 and 2 have anomalously low $^{137}\text{Cs}/^{239,240}\text{Pu}$ activity ratios compared with the other samples from region C and this could be attributed to a small transfer of particulate material from the Irish Sea to this area along with the general dispersal of the soluble components of the Sellafield waste northwards along the Atlantic coast of Scotland^{7,9}. Alternatively, the results may simply relate to the composition of the samples: they were carbonate-rich and low in detrital terrigenous material so that they probably have a low content of fallout nuclides.

The above interpretation is supported by a more detailed analysis in which weighted least square linear regressions were performed for all three combinations of radionuclide concentrations with representative plots being shown in Figs 2 and 3. In region A, the linear correlations are excellent (correlation coefficients: Pu-Am = 0.99; Cs-Am = 0.94; Cs-Pu = 0.95) and the uncertainties on the intercepts in all cases encompass the origin, a combination of circumstances which could not arise if these nuclides were being supplied from a number of different sources, each characterized by its own value of the various nuclide activity ratios. In region B, the Cs-Pu and Cs-Am linear correlations persist (correlation coefficients: Cs-Am = 0.58; Cs-Pu = 0.74), but are noticeably poorer than that for Pu-Am (0.87); also, this analysis shows a non-zero intercept on the Cs axis in both the Cs-Pu and Cs-Am regressions (with the uncertainty on the intercept deduced from the spread of the experimental points) and a non-zero intercept on the Pu axis in the Pu-Am

Table 2 Observed ranges of isotopic and nuclide activity ratios and estimated isotopic and nuclide activity ratios

Results for sediments

Region	$^{137}\text{Cs}/^{239,240}\text{Pu}$	$^{241}\text{Am}/^{239,240}\text{Pu}$	$^{238}\text{Pu}/^{239,240}\text{Pu}$	$^{137}\text{Cs}/^{241}\text{Am}$
A	<10	>1	>0.2	<10
B	10–40	≈1	≈0.2	10–50
C	>40	<1	<0.2	>50

Source terms*

Source	$^{137}\text{Cs}/^{239,240}\text{Pu}$	$^{241}\text{Am}/^{239,240}\text{Pu}$	$^{238}\text{Pu}/^{239,240}\text{Pu}$	$^{137}\text{Cs}/^{241}\text{Am}$
Particle-associated Sellafield waste	1.9	1.04	0.24	1.8
Soluble Sellafield waste	35	<0.28	0.31	>125
Atmospheric fallout	56	0.23	0.05	240

* Ratios appropriate to intertidal sediments in south-west Scotland in 1984–85, arising from the dominant sources^{17,18}.

Table 3 Isotopic and nuclide activity ratios for annual discharges of liquid effluent from Sellafield and the integrated total ^{241}Am to integrated total $^{239,240}\text{Pu}$ ratio

Year	$^{137}\text{Cs}/^{241}\text{Am}$	$^{238}\text{Pu}/^{239,240}\text{Pu}$	Annual discharge $^{241}\text{Am}/^{239,240}\text{Pu}$	Integrated* $^{241}\text{Am}/^{239,240}\text{Pu}$
1970	61		0.55	
1971	35		0.90	
1972	16		1.39	0.52
1973	7		1.67	0.77
1974	34		2.56	1.01
1975	144		0.82	1.02
1976	359		0.25	0.97
1977	1,223		0.10	0.93
1978	516	0.27	0.17	0.90
1979	327	0.32	0.21	0.88
1980	360	0.34	0.40	0.89
1981	269	0.32	0.57	0.92
1982	313	0.29	0.40	0.93
1983	546	0.33	0.25	0.94
1984	189	0.31	0.28	0.94

Activity ratio data from refs 13, 14.

* Includes radioactive decay and growth of ^{241}Am from ^{241}Pu .

case. This clearly implies that in region B there is an excess input of the more soluble species, but that the particulate influence is sufficiently large for the linear correlation not to be destroyed. The evidence appears most convincing for the Pu-Am case, since the Cs and Pu results in region B are quite tightly grouped, but, irrespective of the fine details, the existence of these correlations is a most significant observation given the vastly different chemical behaviour of these three nuclides and the variability of, and between, their outputs from Sellafield over the past twenty years. Similar correlations can be discerned in published MAFF data¹⁸ for the Irish Sea. The Cs-Pu correlation disappears in region C and this is again consistent with supply from more than one source.

Finally, it should be noted that some interesting disparities exist between the isotopic and nuclide activity ratios observed in Irish Sea intertidal sediments and those which might be expected on the basis of the Sellafield discharge coupled with the accepted geochemical behaviour of the radionuclides. Thus, for example, the $^{137}\text{Cs}/^{241}\text{Am}$ activity ratios for the Sellafield discharge (Table 3) in conjunction with K_D values for ^{137}Cs and ^{241}Am of the order of 10^2 and 10^6 respectively might be expected to produce sediment $^{137}\text{Cs}/^{241}\text{Am}$ ratios in the range 0.03–0.3. The lowest $^{137}\text{Cs}/^{241}\text{Am}$ value observed in the present study is 4.5 and MAFF results¹⁸ for Irish Sea intertidal sediments have typically shown $^{137}\text{Cs}/^{241}\text{Am}$ ratios in the range 4 to 10 during the period 1977 to 1985 with the lowest reported value, for Newbiggin, being 1.8. There is at present no obvious explanation of this apparent anomaly on the basis of the available information, although the suggestion by Miller *et al.*¹² that some of the ^{137}Cs may be discharged from Sellafield in particulate form is worthy of further investigation. The $^{241}\text{Am}/^{239,240}\text{Pu}$ results for intertidal sediments in the north-east Irish Sea also present an interesting comparison with the corresponding values for the Sellafield discharge (Table 3). The values greater than unity obtained both in the present work and in MAFF surveys¹⁸ are clearly much larger than would be expected on the basis of the annual discharge figures and the fact that over 90% of the $^{239,240}\text{Pu}$ is believed to be retained in the Irish Sea sediments^{1–4}. The sediment $^{241}\text{Am}/^{239,240}\text{Pu}$ ratios are in fact similar to the total integrated ^{241}Am to total integrated $^{239,240}\text{Pu}$ ratio in the Sellafield discharge (Table 3) and this suggests a substantial degree of mixing within an exchangeable reservoir of sediment in the Irish Sea prior to deposition in intertidal areas. Mixing of this type is also implied by the sediment $^{238}\text{Pu}/^{239,240}\text{Pu}$

activity ratios found in this study (Table 1) and, as discussed in detail by Hunt,²¹ by MAFF results for Irish Sea intertidal sediments: these have typically shown values in the range 0.21 to 0.26 during the period 1977–84 while the ratio in the discharge has been about 0.3 throughout this period. Such mixing is obviously important in the study of sedimentation processes³.

We thank Strathclyde Regional Council for partial financial support of this work, Prof. J. M. Reid for discussions and collaboration and Mr. and Mrs. R. Hill for assistance with sample collection.

Received 23 February; accepted 22 July 1987.

- Hetherington, J. A. in *Environmental Toxicity of Aquatic Radionuclides: Models and Mechanisms* (eds Miller, M. W. & Stannard, J. N.) 82–106 (Ann Arbor Science, Ann Arbor, 1975).
- Hetherington, J. A. *Mar. Sci. Commun.* **4**, 239–274 (1978).
- Smith, T. J., Parker, W. R. & Kirby, R. *Institute of Oceanographic Sciences Rep. No. 110* (1980).
- Pentreath, R. J., Lovett, M. B., Jefferies, D. F., Woodhead, D. S., Talbot, J. W. & Mitchell, N. T. *Proc. IAEA Conf. Radioactive Waste Management* **5**, 315–329 (1984).
- Pentreath, R. J., Harvey, B. R. & Lovett, M. B. in *Speciation of Fission and Activation Products in the Environment* (eds Bulman, R. A. & Cooper, J. R.) 312–325 (Elsevier, London, 1986).
- Jefferies, D. F., Preston, A. & Steele, A. K. *Mar. Pollut. Bull.* **4**, 118–122 (1973).
- Livingston, H. D. & Bowen, V. T. *Nature* **269**, 586–588 (1977).
- Baxter, M. S., McKinley, I. G., MacKenzie, A. B. & Jack, W. *Mar. Pollut. Bull.* **10**, 116–120 (1979).
- McKinley, I. G., Baxter, M. S. & Jack, W. *Estuarine Coastal Shelf Sci.* **13**, 59–67 (1981).
- McKinley, I. G., Baxter, M. S., Ellett, D. J. & Jack, W. *Estuarine Coastal Shelf Sci.* **13**, 69–82 (1981).
- Livingston, H. D., Bowen, V. T. & Kupferman, S. L. *J. mar. Res.* **40**, 253–272 (1982).
- Miller, J. M., Thomas, B. W., Roberts, P. D. & Creamer, S. C. *Mar. Pollut. Bull.* **13**, 315–319 (1982).
- Cambray, R. S. *AERE Harwell Rep. AERE-M-3269* (1982).
- Annual Reports on Radioactive Discharges and Monitoring of the Environment 1982–84* (British Nuclear Fuels Ltd Health and Safety Directorate, Risley).
- Day, J. P. & Cross, J. E. *Nature* **292**, 43–45 (1981).
- MacKenzie, A. B. & Scott, R. D. *Nucl. Eng.* **25**, 110–122 (1984).
- Holm, E. & Persson, R. B. R. *Radioactivity in the Sea* No. 62 (IAEA, 1977).
- Hunt, G. J. *Aquatic Environment Monitoring Reports* Nos 3, 4, 6, 8, 9, 11, 12, 13 (Ministry of Agriculture Fisheries and Food, Lowestoft, 1979–85).
- Nelson, D. M. & Lovett, M. B. *Proc. IAEA Conf. Impacts of Radionuclide Releases into the Marine Environment* 105–119 (IAEA, Vienna, 1981).
- Sholkovitz, E. R. & Mann, D. R. *Geochim. cosmochim. Acta* **48**, 1107–1114 (1984).
- Hunt, G. J. *Sci. Total Environ.* **46**, 261–278 (1985).
- MacKenzie, A. B., Baxter, M. S., McKinley, I. G., Swan, D. S. & Jack, W. *J. radioanal. Chem.* **48**, 29–47 (1979).
- Coleman, G. H. *The Radiochemistry of Plutonium* (US National Academy of Sciences Publ. NASNS 3058) (1965).
- Williams, T. M. thesis, Univ. Strathclyde (1985).
- MacKenzie, A. B. & Scott, R. D. *Nature* **299**, 613–616 (1982).
- Nelson, D. M. & Lovett, M. B. *Nature* **276**, 599–601 (1978).
- Cambray, R. S. & Eakens, J. D. *AERE Harwell Rep. AERE-R-9807* (1980).
- Cambray, R. S., Playford, K. & Lewis, G. N. J. *AERE Harwell Rep. AERE-R-10485* (1982).

The rate of acidification of aquatic ecosystems in Ontario, Canada

P. J. Dillon, R. A. Reid & E. de Grosbois

Ontario Ministry of the Environment, Dorset Research Centre,
PO Box 39, Dorset, Ontario, Canada POA 1E0

Although the acidification of lakes and streams in several parts of the world as a consequence of the emission and deposition of S and N oxides is well-documented^{1–4}, no direct measurement by continuous monitoring of acidification rate of any lake has been reported. In many cases, historical data have been compared to recent data and acidification of aquatic systems, generally over periods of decades, demonstrated. But methodological changes have always compromised the results. In other cases, historical chemical conditions of lakes and streams have been inferred from either geochemical relationships⁵ or from palaeolimnological records^{6,7} and compared to current conditions. Here we present the first direct observations of the acidification rate of a lake that has undergone extensive damage to its aquatic biota as the result of deposition of acids of anthropogenic origin. A threefold decrease in alkalinity and a decrease of 0.2 pH units occurred in the lake despite a reduction in the deposition rate of strong acids over the

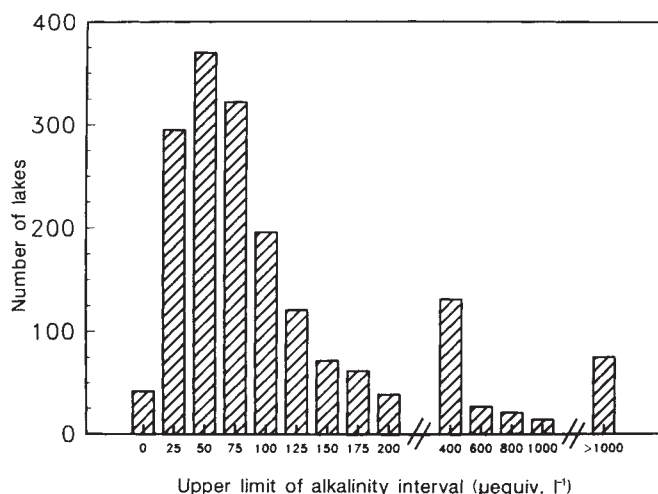


Fig. 1 Alkalinity distribution of lakes in central Ontario. A total of 1,787 lakes were surveyed in the years 1981–85. All lakes were in the SO_4 -deposition zone defined as having SO_4^{2-} deposition of $>60 \text{ mequiv m}^{-2} \text{ yr}^{-1}$.

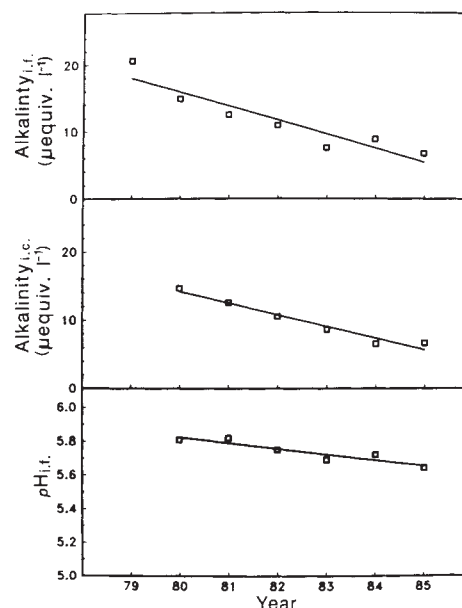


Fig. 2 Alkalinity in the ice-free (i.f.) and ice-covered (i.c.) seasons and pH in the ice-free season in Plastic Lake between the start of the study and 1985. The coefficient of variation within years averaged 20% for the ice-covered period (range 7–41%) and 37% (range 26–53%) for the ice-free period. The lines shown are the least squares linear regression lines.

period of study. Acidification was accompanied by a decrease in base cation content in the lake but not an increase in strong acid anion (SO_4^{2-}) concentration, indicating that there has probably been a depletion of available cations in the lake's catchment.

We have studied two lakes, Plastic (area 32.3 hectares, mean depth 8.0 m) and Harp (area 71.4 ha, mean depth 12.4 m), in the Precambrian part of southern Ontario, Canada since 1979. The catchment of Plastic Lake (88.6 ha) is underlain by orthogneiss and covered by very thin sandy basal till deposits generally $<1.0 \text{ m}$ thick with many bedrock ridges. The upland soils are weakly developed podzols while organic soils and gleysols occur in the depressions and lowlands. Six inflowing streams, five of them ephemeral, drain Sphagnum-conifer swamps or beaver ponds. The streams have annual mean pH values varying between 4.2 and 5.0, and four have high levels of dissolved organic matter during periods of low flow. Nevertheless, most of the annual acidity flux results from the transport of strong acids, particularly H_2SO_4 of anthropogenic origin, not from the transport of organic acids⁸. Harp Lake's catchment (471 ha) is underlain by granitized biotite and hornblende gneiss, with a small area of amphibolite and schist. The glacially-deposited overburden is much thicker than at Plastic Lake, with approximately one-half of the catchment covered by $>1 \text{ m}$ of a sandy loam and sandy till plain, more than 10 m thick in the bottoms of some of the valleys. There are six inflows to Harp Lake, four of which drain small wetland areas and as a result have relatively high levels of dissolved organic matter. The mean annual pHs of the six streams range from 5.5 to 6.3.

There are approximately 85 seasonal dwellings around Harp Lake but none in the Plastic Lake catchment. There has been little or no change in the extent of human activities or alteration of land use in the catchments of the two lakes for at least the past two decades. The amount of precipitation averaged 1,079 mm over the study period (standard deviation 77 mm); there was no trend in either the total annual amount or the annual amount excluding snowfall.

The chemical characterization of the two lakes at the start of the study is shown in Table 1. Plastic Lake and Harp Lake were initially picked as representative of lakes in the zone of high SO_4^{2-} and strong acid deposition which are very sensitive and relatively insensitive, respectively. Of some 1,800 lakes surveyed in the study area, ~20% had alkalinity $<25 \text{ µequiv. l}^{-1}$ (Fig. 1); thus, Plastic Lake has a level of acid neutralizing capacity that is typical of lakes considered to be sensitive to acid deposition

in this region. Harp Lake, on the other hand, had alkalinity near the median value of all lakes surveyed.

Plastic and Harp Lakes were sampled, on average, biweekly during the ice-free period of each year (May to November) but much less frequently ($n = 2-4$) during the winter under ice-cover. Because of the different sampling frequencies between the two periods, data collected in the ice-free and ice-covered seasons are treated separately for any parameters that demonstrated significant variability between seasons.

It is generally accepted that, with respect to the effects of strong acid inputs, alkalinity is the most satisfactory measure of a lake's chemical state because it is conservative with respect to $\text{H}_2\text{CO}_3^* (= \text{CO}_2(\text{aq}) + [\text{H}_2\text{CO}_3])$. We have measured the alkalinity of Plastic and Harp Lakes using a modified version of the Gran titration⁹. Neither the sampling techniques nor the analytical methodology have been altered throughout the period of study.

Table 1 Chemical characterization of Plastic and Harp Lakes in 1980–81

	Plastic Lake	Harp Lake
Alkalinity	14.7	61.1
pH	5.80	6.27
Conductivity	23.9	35.3
DOC	2.7	4.1
SO_4^{2-}	139	173
M^+	168	280
Cl^-	11	22
TP	7.8	8.2
Si	320	1,470
NO_3^-	20	110
Al	34	38

All results are averages for the 1980–81 hydrologic year ($n = 13-27$). Alkalinity (µequiv. l^{-1}) was measured by Gran titration. Conductivity is reported in µS cm^{-1} , dissolved organic carbon (DOC; a measure of the organic content of the water) as mg C l^{-1} , total phosphorus (TP), NO_3^- and reactive silicate (Si) in µg l^{-1} , SO_4^{2-} , Cl^- and base cations ($\text{M}^+ = \text{Ca}^{2+} + \text{Mg}^{2+} + \text{Na}^+ + \text{K}^+$) in µequiv. l^{-1} .

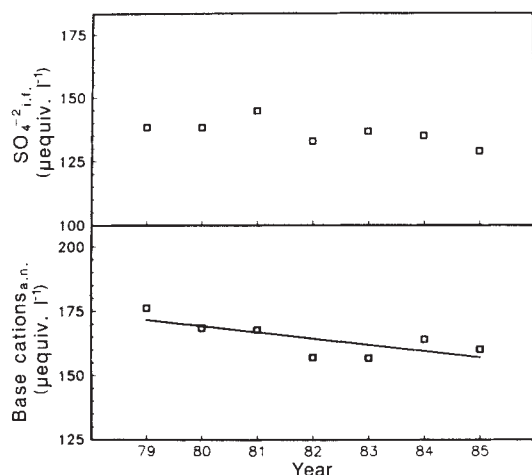


Fig. 3 Sulphate (ice-free period) and base cation concentration (annual) in Plastic Lake. There was no significant trend in SO_4^{2-} concentration; the line shown is the least squares linear regression line.

The alkalinity of Plastic Lake (Fig. 2) decreased from 1979 to 1985. Using simple linear regression analysis, we calculated that the alkalinity decreased by 2.1 ± 0.9 (95% confidence interval) and 1.7 ± 0.5 $\mu\text{equiv. l}^{-1} \text{yr}^{-1}$ during the ice-free and ice-covered seasons, respectively. On the other hand, Harp Lake's alkalinity has not decreased over the same time period; the slope of the regression line is not significantly different from 0 (Table 2).

We also measured the total titratable alkalinity to a fixed endpoint of $\text{pH} = 4.5$ ($\text{Alk}_{4.5}$), a measurement that has the advantage of the end-point being fixed rather than variable. The $\text{Alk}_{4.5}$ of Plastic Lake in the ice-free and ice-covered seasons declined by 1.1 ± 0.6 and 2.5 ± 0.7 $\mu\text{equiv. l}^{-1} \text{yr}^{-1}$, respectively, while the $\text{Alk}_{4.5}$ of Harp Lake did not change significantly.

Although pH is affected by the CO_2 partial pressure of the water, which may not have been constant between years, we measured pH because of its importance to aquatic biological communities^{10,11}. The annual pH of Plastic Lake decreased by an average of 0.035 ± 0.014 pH units per year, while that of Harp did not change (Table 2).

The decrease in alkalinity in Plastic Lake must have been accompanied by either a decrease in base cations (Ca^{2+} , Mg^{2+} , Na^+ , K^+) or an increase in strong acid anions (SO_4^{2-} , NO_3^- , Cl^-) or organic acid anions. Analysis of the SO_4^{2-} and base cation data (Table 2) demonstrated that there was no significant change in SO_4^{2-} concentration but that the base cation concentration decreased by an average of 2.25 ± 1.4 $\mu\text{equiv. l}^{-1} \text{yr}^{-1}$. Nitrate made up an insignificant portion of the anion content of the lake at all times, while the Cl^- concentration, although low, decreased (by 0.7 ± 0.2 $\mu\text{equiv. l}^{-1} \text{yr}^{-1}$) over the course of the study. In Harp Lake, on the other hand, SO_4^{2-} concentration decreased significantly (4.1 $\mu\text{equiv. l}^{-1} \text{yr}^{-1}$) over the study period. Although there was no corresponding increase in alkalinity, there was a comparable, but not statistically significant, decrease in base cations.

Although it has been argued that acidification of freshwaters can be attributed to increases in organic acids of terrestrial origin¹², all available evidence indicates that acidification of lakes is generally accompanied by a decrease in organic content^{13,14}. The dissolved organic carbon content of Plastic Lake decreased by 0.10 ± 0.05 $\text{mg C l}^{-1} \text{yr}^{-1}$, and as the pH has also decreased, the organic acid anion content must have declined¹⁵. Furthermore, the water's transparency, as measured by Secchi disc depth, increased by 0.23 ± 0.14 m yr^{-1} , consistent with the decrease in dissolved organic matter. Thus, acidification of Plastic Lake has resulted in a loss of organic acids of humic

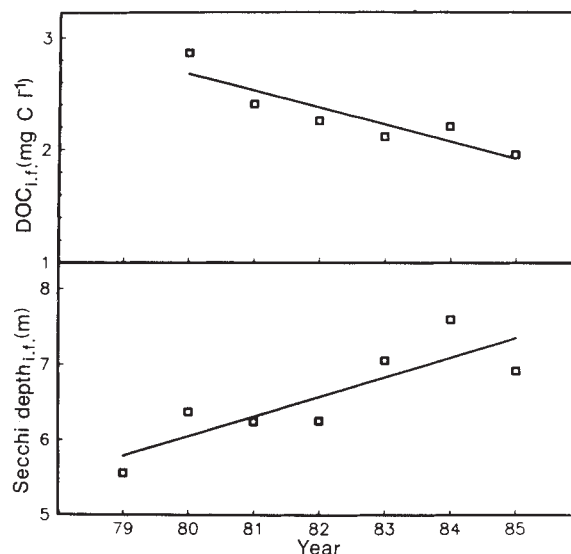


Fig. 4 Dissolved organic carbon (DOC) and Secchi disk depth (both in the ice-free period) in Plastic Lake between the beginning of the study and 1985. The lines shown are the least squares linear regression lines. The coefficient of variation within years averaged 6% (range 3–9%) for DOC and 17% (range 15–19%) for Secchi disk depth.

origin from the water column. In Harp Lake, there was no change in either dissolved organic carbon (DOC) or Secchi disk depth over the study period; thus, the decrease in SO_4^{2-} in Harp Lake was not offset by an increase in organic acid anions.

We have measured no change in the rate of input of base cations from the atmosphere over the study period, nor do we expect an increased rate of loss to the sediments; therefore, the decrease in base cation content of the lake indicates that the supply from the catchment has decreased. As there were no trends in amount of precipitation or runoff, this implies that the pool of available base cations in the terrestrial system may have decreased in size, either because of depletion or because of a reduced replenishment rate by weathering. Analysis of the trends in the individual components of M^+ indicated that the proportion of the decrease in total base cations attributable to Ca^{2+} was 60%, K^+ 12% and Na^+ 28%, while Mg did not change. Of the three cations that changed, the decrease in the relative amount of K^+ , the base cation present in the smallest quantities in Plastic Lake, was greatest (30%). Because the contribution from the atmosphere is insignificant, as a first approximation, this decrease represents a decline of $\sim 30\%$ in the amount of K^+ exported from the catchment in solution. If the availability to the terrestrial plants has dropped by a similar amount, this decrease would be significant because of the importance of K^+ as a nutrient for north-temperate forests¹⁶.

An alternate mechanism that could explain the reduction in base cations in the catchment's runoff as well as the acidification of Plastic Lake is the net accumulation of base cations in the forest biomass. But there has been no logging in the Plastic Lake catchment since the first decade of this century; we therefore believe that accelerating base cation accumulation in biomass is not a significant contributor to the acidification process at this site.

Deposition of strong acid and SO_4^{2-} in the study area has decreased over the duration of this investigation (data not shown). Sulphate deposition ranged from 74 to 87 $\text{mequiv. m}^{-2} \text{yr}^{-1}$ between 1976 and 1980 but began to decline as SO_2 emissions in eastern North America declined, such that the level was ~ 55 – 60 $\text{mequiv. m}^{-2} \text{yr}^{-1}$ in 1983–86. Deposition of H^+ followed a similar pattern; however, Plastic Lake has continued to acidify as a result of the continuing desorption of SO_4^{2-} from the catchment despite the reduced deposition. The

Table 2 Summary of trends in the chemical content of Plastic and Harp Lakes

Parameter		Annual change	s.e.	P	r
Plastic Lake					
Alkalinity	(i.f.)	-2.1	0.36	0.002	0.93
Alkalinity	(i.c.)	-1.7	0.17	0.0006	0.98
Alkalinity _{4.5}	(i.f.)	-1.1	0.21	0.006	0.94
Alkalinity _{4.5}	(i.c.)	-2.5	0.29	0.0004	0.97
pH	(i.f.)	-0.035	0.007	0.008	0.93
M ⁺	(a.n.)	-2.25	0.73	0.02	0.78
SO ₄ ²⁻	(i.f.)	-----	NS	-----	-----
Cl ⁻	(i.f.)	-0.71	0.09	0.0002	0.96
DOC	(a.n.)	-0.10	0.02	0.004	0.92
Secchi depth	(i.f.)	+0.23	0.06	0.007	0.86
Harp Lake					
Alkalinity		-----	NS	-----	-----
pH		-----	NS	-----	-----
M ⁺	(a.n.)	-----	NS	-----	-----
DOC	(a.n.)	-----	NS	-----	-----
SO ₄ ²⁻	(i.f.)	-4.1	0.65	0.0004	0.92

Data were analysed as ice-free (i.f.) and ice-covered (i.c.) season averages if there was significant within-year variability and annual (a.n.) averages if there was no seasonal variability. The annual change (with standard error (s.e.), significance level (*P*) and linear correlation coefficient (*r*) was estimated by linear regression, and is reported in $\mu\text{equiv. l}^{-1} \text{yr}^{-1}$ for all alkalinity parameters, SO_4^{2-} , base cations ($\text{M}^+ = \text{Ca}^{2+} + \text{Mg}^{2+} + \text{Na}^+ + \text{K}^+$), and Cl, in $\text{mg Cl}^{-1} \text{yr}^{-1}$ for DOC and in m yr^{-1} for Secchi depth. For Harp Lake, all correlations for alkalinity and pH for different seasons were not significant (NS).

reduction in sulphate and H^+ deposition on account of the lower SO_2 emissions may eventually slow or even reverse the rate of acidification of Plastic Lake, but because the S may have a long residence time in the catchment and because the theoretical water replenishment time of the lake is four years, a new steady-state lake chemistry would be expected to take considerably longer to reach. The biota of Plastic Lake have experienced major detrimental effects in the past eight years including fish kills¹⁷, extinction or near-extinction of mollusc species¹⁸, and amphipods^{19,20} as well as proliferation of acidophilic filamentous algae²¹. In addition, it has recently been shown that biological damage may result at alkalinities and pHs greater than those measured in Plastic Lake^{22,23}. Further reductions in strong acid/ SO_4^{2-} deposition are therefore essential if the biological damage in Plastic Lake is to be reversed.

In conclusion, Plastic Lake acidified between 1979 and 1985 at the rate of $\sim 2 \mu\text{equiv. alkalinity lost l}^{-1} \text{yr}^{-1}$. The loss corresponded to a decrease in base cation content, but not an increase in strong acid anion (SO_4^{2-}) content. The dissolved organic content of the lake decreased simultaneously. As a result of recent decreases in atmospheric S deposition in eastern North America, the rate of decrease of alkalinity may slow down, but further decreases in deposition of strong acids and acid precursors are necessary if biological damage is to be reversed. Based on chemical surveys, we believe that Plastic Lake is typical of many thousands of lakes in eastern North America that are sensitive to acidic deposition.

This work was carried out as part of the Acidic Precipitation in Ontario Study. We thank B. LaZerte, W. Keller, W. Scheider and A. Neary for providing comments on the manuscript, B. Locke for help with the data analysis, R. Girard for sample collection and B. Neary for permission to use the data in Fig. 1.

Received 15 April; accepted 8 July 1987.

1. National Research Council *Acid Deposition Long-term Trends* (National Academy Press, Washington DC, 1986).
2. Environmental Protection Agency *The Acidic Deposition Phenomenon and Its Effects. Critical Assessment Review, Vol. II* (eds Altshuler, A. P. & Linthurst, R.) EPA-600/8-83-016 BF (1984).

3. Harvey, H. H. *et al. Acidification in the Canadian Aquatic Environment*. Natn. Res. Council Can. Publ. No. 18475 (1981).
4. Wright, R. F. *Predicting Acidification of North American Lakes*. Norwegian Institute for Water Research Rep. 4/1983 (1983).
5. Henriksen, A. *Nature* **278**, 542-545 (1979); *Changes in Base Cation Concentrations Due to Freshwater Acidifications*. Norwegian Institute for Water Research Rep. 1/1982 (1982).
6. Flower, R. J. & Battarbee, R. W. *Nature* **305**, 130-133 (1983).
7. Jones, V. J., Stevenson, A. C. & Battarbee, R. W. *Nature* **322**, 157-158 (1986).
8. LaZerte, B. D. & Dillon, P. J. *Can. J. Fish. Aquat. Sci.* **41**, 1664-1677 (1984).
9. Stumm, W. & Morgan, J. J. *Aquatic Chemistry. An Introduction Emphasizing Chemical Equilibria in Natural Waters* 2nd edn (Wiley, New York, 1981).
10. Haines, T. A. *Trans. Am. Fish. Soc.* **110**, 669-707 (1981).
11. Dillon, P. J., Yan, N. D. & Harvey, H. H. *CRC Crit. Rev. Envir. Contr.* **13**, 167-194 (1984).
12. Krug, E. C. & Frink, C. R. *Science* **221**, 520-525 (1983).
13. Almer, B. *et al. Ambio* **3**, 30-36 (1974).
14. Davis, R. B., Anderson, D. S. & Berge, F. *Nature* **316**, 436-438 (1985).
15. Oliver, B. G., Thurman, E. M. & Malcolm, R. L. *Geochim. cosmochim. Acta* **47**, 2031-2035 (1983).
16. Morrison, I. *Mineral Nutrition of Conifers with Special Reference to Nutrient Status Interpretation: a Review of Literature*. Can. Forest Serv. Publ. No. 1343 (1974).
17. Harvey, H. H. & Lee, C. *Historical Fisheries Changes Related to Surface Water pH Changes in Canada. Acid Rain/Fisheries*. Proceedings of an International Symposium on Acidic Rain and Fishery Impacts on Northeastern North America, Cornell University, Ithaca, New York (1982).
18. Servos, M. R., Rooke, J. B. & Mackie, G. L. *Can. J. Zool.* **63**, 511-515 (1985).
19. Stephenson, M. & Mackie, G. L. *Can. J. Fish. Aquat. Sci.* **43**, 288-292 (1986).
20. France, R. L. & LaZerte, B. D. *Can. J. Fish. Aquat. Sci.* (in the press).
21. Jackson, M. B. *Wat. Air Soil Pollut.* (in the press).
22. Schindler, D. W. *et al. Science* **228**, 1395-1401 (1985).
23. Magnuson, J. J., Baker, J. P. & Rahel, E. J. *Phil. Trans. R. Soc. B* **305**, 501-516 (1984).

Isotopic compositions and probable origins of organic molecules in the Eocene Messel shale

J. M. Hayes*, Ray Takigiku*, Ruben Ocampo†, Henry J. Callot† & Pierre Albrecht†

* Biogeochemical Laboratories, Departments of Chemistry and Geology, Geology Building, Indiana University, Bloomington, Indiana 47405-5101, USA

† Laboratoire de Chimie Organique des Substances Naturelles, UA 31 associée au CNRS, Département de Chimie, Université Louis Pasteur, 1, rue Blaise Pascal, 67008 Strasbourg, France

The sediments that now comprise the Messel shale¹ accumulated 47 ± 2 million years ago in anaerobic waters at the bottom of a lake². Subsequent depths of burial have not exceeded 300 m, nor has the temperature of the shale exceeded 40°C ³. Contents of organic carbon reach 25%, and preservation of molecular structures has been excellent^{4,5}. Sixteen different geoporphyrins, including three derived from bacteriochlorophylls of the *d* series and thus indicative of the existence in the lake of an anaerobic photic zone, have been isolated and identified⁶⁻⁸. Here, we show that the carbon isotopic compositions of these and other biomarkers allow identification of specific sources for some materials and reconstruction of carbon flows within the lake and its sediments. The ^{13}C content of organic matter synthesized by lacustrine primary producers can be estimated from the observed ^{13}C content of the geoporphyrins derived from their chlorophylls. Total organic material in the shale is depleted in ^{13}C by six parts per thousand relative to that input. This difference cannot be explained by selective loss of components enriched in ^{13}C , nor, as shown by isotopic compositions of other biomarkers, by inputs from land plants surrounding the lake or from methanogenic bacteria.

Isotopic compositions of carbon fractions are summarized in Table 1. Carbon isotopic compositions and specific or probable biological sources of biomarkers in the Messel shale (structures are shown in Fig. 1), including the eight most abundant porphyrins, are summarized in Table 2. The identifications of specific precursors for the porphyrins^{6,8} noted in Table 2 show that algae and photosynthetic bacteria were important contributors of porphyrin precursors to the sediment (a correlation of a petroporphyrin with a specific precursor was also suggested for chlorophyll *b*, ref. 9). Chlorophyll *c* occurs uniquely (accompanying chlorophyll *a*) in brown algae (Chrysophyceae and

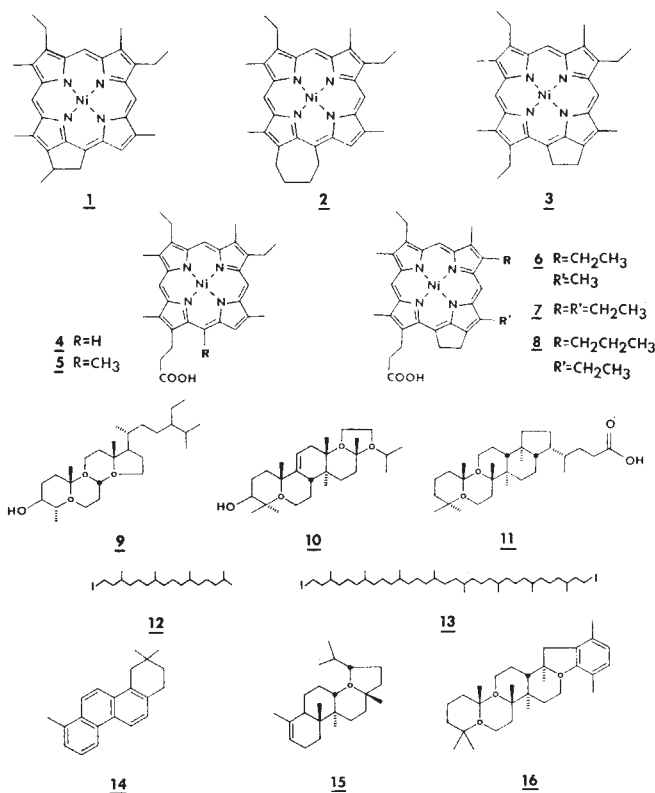


Fig. 1 Structures of compounds for which isotopic compositions are reported. For methods of isolation and analysis see references cited in Table 2.

Haptophyceae), diatoms (Bacillariophyceae), and dinoflagellates (Dinophyceae)¹⁰. The abundance of 4-methyl sterols in these sediments (for example compound 9, see also refs 5, 11, 12) suggests that the last of these alternatives is most probable. Bacteriochlorophylls in the *d* series are restricted to the green photosynthetic bacteria (Chlorobiaceae)¹³.

The porphyrin to which an algal origin can be assigned on purely structural grounds (compound 1, -22.15%) is enriched in ¹³C by ~2% relative to the porphyrins for which a photosynthetic-bacterial origin is required (7, -23.96; 8, -23.92%). Among the remaining geoporphyrins, which must be related to chlorophyll or bacteriochlorophyll because they retain either the five-membered isocyclic ring or its degradation products, 3, 4 and 5 have isotopic compositions near that of 1, suggesting that these porphyrins also have an algal origin. In contrast, the isotopic composition of 6 falls between those of the compounds of purely algal or bacterial origin and is apparently representative of a mixture.

Structurally, 6 and 3 appear related as diagenetic precursor and product. The isotopic difference between them, however, indicates that 3 is not simply the decarboxylation product of 6. Material derived from photosynthetic bacteria has contributed to the carboxyl-bearing structure, but not (appreciably) to the alkyl porphyrin. The small but significant difference may be explained by either an easier (acrylic against propionic side-chain) or earlier decarboxylation of the algal precursor.

The isotopic composition of 2 is, like its structure^{14,15}, unique. It has been suggested^{14,15} that this carbon skeleton could derive from an unknown or extinct precursor, or could be produced from chlorophyll *a*, *b* or *c*. The present results favour a unique origin. Significant isotopic fractionations accompanying production of 2 from chlorophyll *a*, *b* or *c* would be expected at, at most, two of the 32 carbon positions within the molecule. If, by some means, ¹³C were enriched at these positions, the magnitude of the enrichment would have to be 32% at each position (or 64% at a single position) to explain the overall 2% enrichment

Table 1 Isotopic compositions of carbon fractions in the Messel shale

Sample identification	$\delta^{13}\text{C}_{\text{PDB}}$ (‰)
Total carbonate	+7.34 ± 0.12
Total organic carbon, kerogen*	-28.21 ± 0.03
Total extractable organic material†	-29.72 ± 0.10
Total extract fractionated on SiO ₂ column:‡	
Hexane eluent	-33.66 ± 0.14
Toluene eluent	-29.66 ± 0.03
Methanol eluent	-28.30 ± 0.07
Alkyl porphyrin fraction (Strasbourg)	-22.60 ± 0.08
Total porphyrins (Bloomington)	-23.43 ± 0.06
Acid porphyrin fraction (Strasbourg)	-23.91 ± 0.04

Carbon dioxide for isotopic analysis prepared by combustion of organic material in sealed quartz tubes³³. Carbon-isotopic compositions reported in parts per thousand relative to the PDB standard³⁴.

* Preparation: ref. 33.

† Soxhlet extraction, 50/50 (v/v) dichloromethane-methanol, 72 h.

‡ 'Baker-10' disposable extraction column.

observed (-19.5% for 2 versus -21.5% for other algal porphyrins). Enrichments of this magnitude can be excluded, and it is required that 2 has some unique source. A natural compound possessing the same seven-membered ring has been isolated from the sponge, *Dariwinella oxecta*¹⁶.

To develop an estimate of the isotopic composition of the organic carbon produced by the plants which synthesized the porphyrin precursors, we note that the tetrapyrrole nucleus of chlorophyll is commonly enriched in ¹³C by ~0.5% relative to the total plant. Specifically, Park and Dunning¹⁷ reported that the chlorin nucleus of chlorophyll from tomato leaves was enriched in ¹³C by 0.5% relative to the whole leaf. At Indiana (R. Takigiku, G. Vasquez, H. Gest and J. M. Hayes, unpublished data), we have found that the chlorin nucleus from *Rhodospseudomonas capsulata* is depleted in ¹³C by 0.01% relative to total cells and that, in *Chromatium vinosum*, the chlorin nucleus is enriched by 0.73%. In leaves from a beech tree, the enrichment was 0.66%. The inferred isotopic compositions of the algae and photosynthetic bacteria in the Messel Lake are, therefore, ~-22.5 and -24.5%.

The Messel shale contains macro and molecular fossils of higher plants^{18,19}. It is likely that the pentacyclic triterpenol, 10 (ref. 18), for which no specific source has been determined, is related to higher-plant inputs. Its isotopic composition falls between those of the modern materials 14 (-28.6%) and 15 (-27.7%), both isolated from sediments of ponds near Strasbourg and probably derived from 3-hydroxy triterpenoids produced by higher plants²⁰. This indication that the isotopic composition of Eocene higher-land plants did not differ greatly from that of similar modern materials is consistent with (1) the observation that the isotopic composition of coals and related materials appears to have varied little since well before Eocene time²¹, and, (2) our measurement of the isotopic composition of leaf material (-27.5%) organically preserved at the Lamkin Clay Pit²² in the Middle Eocene Upper Claiborne Formation of Tennessee. Porphyrins derived from land plants surrounding the Messel Lake are expected, therefore, to have isotopic compositions like those of modern higher-plant tetrapyrroles, or ~-27% (compare with -28.54% for chlorophyllide from beech-tree chloroplasts, R. Takigiku and J. M. Hayes, unpublished; -26.99 ± 0.51% for hemisynthetic porphyrin standards prepared from pheophorbide from stinging nettles as described in refs 7 and 8 and analysed in the course of this work). It is, therefore, notable that porphyrins 3, 4 and 5, structures which could be related to chlorophyll *a* and *b* and which might be associated with higher plants, are slightly enriched in ¹³C relative to 1, the material that certainly originated in the water column. A shift in the opposite direction would be expected if 3, 4 and 5 included significant contributions from chlorophylls synthesized by land plants. Hence, it is likely that 3, 4 and 5 derive exclusively from

algae and that chlorophylls from land plants were quantitatively degraded before reaching the bottom of the lake²³. It is also possible that the mass of higher-plant debris entering the lake was far outweighed by that of the algae and photosynthetic bacteria. If so, the isotopic composition of the average primary input would have been near -23 or -24‰, a compromise between the principal sources. Inputs of higher-plant material would pull the average to lower values, but the kerogen is classified as type II²⁴ (reflective primarily of aquatic inputs), and these must not have been extensive. If one-third of the input derived from higher plants, the average would drop to only -24.8‰.

Apparently, primary organic inputs were reworked or augmented to pull the overall isotopic composition down to -28.2‰ (Table 1). Were heavy (¹³C-enriched) components lost, light components added, or does some combination of these processes explain the isotope shift? The first alternative can be excluded as the sole mechanism. If the isotope shift were due only to loss of components enriched in ¹³C, the lightest materials present should be higher-plant lipids and few primary components enriched in ¹³C could remain. It is, however, evident (Tables 1 and 2) that the materials present cover an isotope range that has been broadened beyond that of the primary inputs. We conclude that at least some portion, and perhaps all, of the change in average isotopic abundances is due to the addition of organic compounds depleted in ¹³C relative to the primary inputs. A significant portion of the material present must have some non-photosynthetic source.

Non-photosynthetic inputs might be either heterotrophic or autotrophic. Heterotrophic processes (aerobic or anaerobic) in which organic molecules with more than two or three carbon atoms serve as the substrate are not likely to yield large isotopic fractionations²⁵, and are, therefore, unlikely sources for compounds strongly depleted in ¹³C. Chemolithotrophic or methylo-trophic processes in which sulphur compounds serve as electron donors or acceptors are not likely to have been important in this non-marine system. The remaining possibility is that processes of methane production and (aerobic) consumption may have been important.

The presence in the Messel shale of biomarkers characteristic of methanogenic bacteria (12, 13, refs 26, 27), the presence of carbonate enriched in ¹³C (Table 1, see ref. 28 for review of

association between ¹³C-enriched carbonates and methanogenesis), and what is known of conditions in the depositional environment² all suggest that methanogenesis was an important process in the Messel Lake. It is widely recognized that some of the largest isotopic fractionations encountered in nature are associated with this process, and it is tempting to link it with the depletion of ¹³C in the Messel sediments. Remarkably, however, the isotopic compositions of 12 and 13 show that the methanogen biomass itself was not strongly depleted in ¹³C. This is not inconsistent with what is known of isotopic fractionations associated with the synthesis of methanogenic biomass. Specifically, it has been found that the isotopic fractionation between methanogenic biomass and source CO₂ is only about 60% as large as that between the CO₂ source and the CH₄ product (R. Takigiku, P. Hartzell and J. M. Hayes, manuscript in preparation)^{29,30}. It can be estimated that the isotopic fractionation between CO₂ and CH₄ in a lake at 25 °C would be 43‰³¹. Therefore, the fractionation expected between CO₂ and methanogenic biomass would be 0.6 × 43 = 26‰. Lipids would probably be depleted by an additional amount (~1‰, ref. 34; up to 10‰, Takigiku, Hartzell and Hayes, manuscript in preparation), yielding a total fractionation between CO₂ and lipids (for example, 12 and 13) of ~31‰. It follows that the isotopic composition of the porewater CO₂ ought to have been near 1‰ (= -30 + 31), a value reasonably consistent with the presence of carbonates with δ = +7.3‰ (Table 1). We conclude that the evidence indicating that methanogenic biomass, in these lacustrine sediments, was not strongly depleted in ¹³C relative to total organic carbon must be accepted at face value and that, therefore, methanogenesis (methane production) cannot be responsible for the overall isotope lightness.

In contrast, recycling of methane carbon might have had significant isotopic consequences. Microaerophilic methylo-trophs living at the anaerobic/aerobic boundary of the lake would consume methane rising from the anaerobic zone and synthesize biomass that, because of its derivation from methane, would be very strongly depleted in ¹³C. Sedimentation and addition of this material to the organic debris accumulating in the sediments might have played an important role in achieving the large overall depletion in ¹³C. An important role for bacteria in development of the overall depletion of ¹³C is indicated by the presence of a bacterial biomarker (11) strongly depleted in ¹³C. A related material (16), isolated³² from a Guatemalan crude oil of Cretaceous (Aptian) age is similarly depleted (-35.0‰), suggesting that bacterial inputs might often be associated with isotopic lightness in sedimentary organic materials. The isotopic composition of the nonpolar hydrocarbon fraction of the extract (Table 1) also shows that lipids more strongly depleted in ¹³C than any derived from methanogens or land plants were present in the biological input to the Messel sediments. Further study of these ¹³C-depleted fractions should allow identification of their source.

We are grateful for support from NASA and the NSF. We thank S. A. Studley for technical assistance, D. Dilcher (Palaeobotanical Collections, Indiana University) for samples of Eocene plant material (collected with the generous support of NSF) and M. Lewan (Amoco Production Company Research Center) for helpful discussions regarding the fate of terrigenous chlorophylls.

Received 11 February; accepted 19 June, 1987.

- Matthes, G. *Mém. Bur. Rech. géol. Minières* **58**, 327-337 (1968).
- Von Koenigswald, W. *Geol. Jb. Hessen* **108**, 23-38 (1980) and references therein.
- Arpino, P., Albrecht, P. & Ourisson, G. in *Advances in Organic Geochemistry 1971* (eds von Gaertner, H. R. & Wehner, H.) 173-187 (Pergamon, Oxford, 1972).
- Chappe, B., Michaelis, W. & Albrecht, P. in *Advances in Organic Geochemistry 1979* (eds Douglas, A. G. & Maxwell, J. R.) 265-274 (Pergamon, Oxford, 1981).
- Dastillung, M., Albrecht, P. & Ourisson, G. *J. chem. Research (S)* **166-167**; (*M*) **2325-2352** (1980).
- Ocampo, R., Callot, H. J., Albrecht, P. & Kintzinger, J. P. *Tetrahedron Lett.* **25**, 2589-2592 (1984).
- Ocampo, R., Callot, H. J. & Albrecht, P. *J. Chem. Soc. Chem. Commun.* **198-200** (1985).
- Ocampo, R., Callot, H. J. & Albrecht, P. *J. Chem. Soc. Chem. Commun.* **200-201** (1985).

Table 2 Carbon isotopic compositions of individual biomarkers

Structure*	Molecular precursor†	Refs‡	Related organism§	Abundance ng g ⁻¹	δ ¹³ C _{PDB} (‰)	n
2		6		1	-19.50 ± 0.05	3
5		7	Algae	2	-21.58 ± 0.13	1
4		7	Algae	1	-21.89 ± 0.15	2
3		6	Algae	4	-21.92 ± 0.08	2
1	Chl c	6	Algae	5	-22.15 ± 0.03	2
6		8	Mixture	12	-23.12 ± 0.04	3
8	Bchl d	8	Photosynth. bacteria	3	-23.92 ± 0.05	3
7	Bchl d	8	Photosynth. bacteria	1	-23.96 ± 0.06	2
9		11	Dinoflagellate		-25.37 ± 0.08	1
10		18	Unknown		-28.05 ± 0.03	1
12	Phytanyl ether	¶	Methanogen	~8	-29.74 ± 0.14	3
13	Biphytanyl ether	¶	Methanogen	~16	-29.88 ± 0.12	3
11		5	Bacteria		-32.27 ± 0.11	2

* Figure 1.

† Identified in reference cited.

‡ Initial report of isolation and structure and, where possible on structural grounds, assignment of specific precursor.

§ Identified by reference to natural distribution of biomarkers among organisms. Italicized assignments derive from this work.

|| Number of isotopic analyses.

¶ From polar extract³⁵ treated with aqueous HI to form polyisoprenoid iodides³⁶. Iodides isolated by chromatography on silica gel, chloroform eluent. Purity assessed by comparison of chromatograms and mass spectra with authentic standards prepared from *Methanobacterium thermoautotrophicum*.

9. Chicarelli, M. I. & Maxwell, J. R. *Tetrahedron Lett.* **25**, 4701-4704 (1984).
10. Jeffrey, S. W. in *Primary Productivity in the Sea* (ed. Falkowski, P. G.) 33-58 (Plenum, New York, 1980).
11. Mattern, G., Albrecht, P. & Ourisson, G. *J. chem. Soc. Chem. Commun.* 1570-1571 (1970).
12. Robinson, N., Eglinton, G., Brassel, S. C. & Cranwell, P. A. *Nature* **308**, 439-441 (1984).
13. Jackson, A. H. in *Chemistry and Biochemistry of Plant Pigments*, 2nd edn. Vol. 1 (ed. Goodwin, T. W.) 1-44 (Academic, London, 1976).
14. Wolff, G. A., Murray, M., Maxwell, J. A., Hunter, B. K. & Sanders, J. K. M. *J. chem. Soc. Chem. Commun.* 922-924 (1983).
15. Fookes, C. J. R. *J. Chem. Soc. Chem. Commun.* 1474-1476 (1983).
16. Karuso, P. *et al. Tetrahedron Lett.* **27**, 2177-2178 (1986).
17. Park, R. & Dunning, H. N. *Geochim. Cosmochim. Acta* **22**, 99-105 (1961).
18. Albrecht, P. & Ourisson, G. *Science* **163**, 1192-1193 (1969).
19. Spyckerelle, C., Greiner, A. C., Albrecht, P. & Ourisson, G. *J. chem. Research (S)* 330-331; (M) 3746-3777 (1977).
20. Corbet, B., Albrecht, P. & Ourisson, G. *J. Am. chem. Soc.* **102**, 1171-1173 (1980).
21. Redding, C. E., Schoell, M., Monin, J. C. & Durand, B. in *Advances in Organic Geochemistry 1979* (eds Maxwell, J. R. & Douglas, A. G.) 711-723 (Pergamon, Oxford, 1980).
22. Potter, F. W. Jr *Paleontographica* **157**, 44-96 (1976).
23. Sanger, J. E. *Ecology* **52**, 1075-1089 (1971).
24. Tissot, B. P. & Welte, D. H. *Petroleum Formation and Occurrence*, 2nd edn **257** (Springer, Berlin, 1984).
25. DeNiro, J. & Epstein, S. *Geochim. cosmochim. Acta* **42**, 495-506 (1984).
26. Chappe, B., Albrecht, P. & Michaelis, W. *Science* **217**, 65-66 (1982).
27. Michaelis, W. & Albrecht, P. *Naturwissenschaften* **66**, 420-422 (1979).
28. Irwin, H., Curtis, C. & Coleman, M. *Nature* **269**, 209-213 (1977).
29. Fuchs, G., Thauer, R., Ziegler, H. & Stiehler, W. *Arch. Microbiol.* **120**, 135-139 (1979).
30. Belyaev, S. S. *et al. Appl. Environ. Microbiol.* **45**, 691-697 (1983).
31. Whiticar, M. J., Faber, E. & Schoelle, M. *Geochim. cosmochim. Acta* **50**, 693-709 (1986).
32. Hussler, G. *et al. Tetrahedron Lett.* **25**, 1179-1182 (1984).
33. Wedeking, K. W., Hayes, J. M. & Matzke, U. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 428-441 (Princeton, New Jersey, 1983).
34. Craig, H. *Geochim. cosmochim. Acta* **12**, 133-149 (1957).
35. Tornabene, T. G. & Langworthy, T. A. *Science* **203**, 51-53 (1978).
36. Panganamala, R. V., Sievert, C. F. & Cornwell, D. G. *Chem. Phys. Lipids* **7**, 336-344 (1971).

Estimates of degradable organic carbon in deep-sea surface sediments from ^{14}C concentrations

Steven Emerson*, Charles Stump*, Pieter M. Grootes†, Minze Stuiver†, George W. Farwell‡ & Fred H. Schmidt‡

* School of Oceanography, † Quaternary Research Center and Department of Geological Sciences, ‡ Department of Physics and Nuclear Physics Laboratory, University of Washington, Seattle, Washington 98195, USA

Organic carbon sources, degradation and burial in surface marine sediments control the exchange of this carbon pool with that in the ocean and provide the key to understanding the response of the sedimentary reservoir to changing environmental conditions. We explore these subjects by organic-carbon concentration profiles and newly determined radiocarbon activities from three deep-water locations. Previous ^{14}C studies in deep-sea^{1,2} and near-shore³ marine sediments revealed deep penetration of relatively recent organic carbon, some of which was believed to carry ^{14}C from atmospheric nuclear-weapons testing. Our results suggest that the bulk of the organic matter in the surface mixed layer of marine sediments has a planktonic origin, and up to 50% of this reservoir is degradable on a timescale of $\leq 1,000$ yr; short enough to have responded to glacial-interglacial changes. The degradable fraction is estimated to be presently $50 \pm 15 \times 10^{15}$ g C and may have been several times larger during the past glacial period.

There are four potential sources for organic carbon in marine sediments: plankton from surface waters, land plant material introduced by rivers, dissolved organic carbon (DOC) in seawater⁴ and recycled fossil carbon from the weathering of continental rocks⁵. There is no doubt that the main particulate organic carbon flux to the ocean floor originates from marine phytoplankton because of the close relation between particulate organic-carbon flux and benthic oxygen consumption^{6,7}. Whether this is true for the carbon observed in the sedimentary mixed layer is less clear and has a major influence on the interpretation of the ^{14}C ages because of the different radiocarbon activities of the potential sources.

Samples we describe were raised from the sea floor with box corers, which effectively recover the sediment surface. Because the organic carbon content of these sediments is less than 1% (dry mass) and the amount of material available is limited, the ^{14}C analysis was carried out by accelerator mass spectrometry (AMS)⁸⁻¹¹. Results of the organic carbon and organic ^{14}C analyses are presented in Fig. 1 along with calcium carbonate content and CaCO_3 ^{14}C stratigraphy determined on the same core or from published data at the same oceanographic location. Previously published organic ^{14}C data from three cores in the North Pacific² are included for comparison.

The organic carbon data reveal a variable concentration in the surface 8-10 cm and then a decrease to lower, more consistent concentrations below this depth. The depth 8-10 cm corresponds to the biologically homogenized mixed layer, h , based on CaCO_3 ^{14}C data for cores *a-c* and thorium-230 activities for the North Pacific core². If we assume steady-state sedimentation, the carbon profiles indicate a mixture of degradable and more refractory components in the mixed layer with the more refractory material accumulating below this depth. Organic ^{14}C ages are, in all cases, youngest in the top few centimetres and then nearly homogeneous throughout the mixed layer. At least part of the reason for younger ages at the top is probably the contribution of bomb-produced ^{14}C because other isotopes that originate from nuclear-weapons testing have been observed in deep-sea sediments². Mean mixed-layer ages lie between 2,000 and 4,700 yr BP in this study and $\sim 12,000$ yr BP in the North Pacific core². The latter has an accumulation rate, s , of a factor of ten lower (Fig. 1 caption) indicating that, to a first approximation, the residence time of sediment in the mixed layer ($\tau_r = h/s$) has a strong influence on the organic ^{14}C age. The ^{14}C activity below the mixed layer was measured on the South Pacific core studied here and the North Pacific samples²; for both the ages are much greater at these depths (Fig. 1).

The stable carbon isotope ratio ($\delta^{13}\text{C}$) of organic matter is a tracer for its origin because the ratio for terrestrial organic carbon ($\delta^{13}\text{C} \sim -25$) is different from that produced by plankton in temperate ocean waters ($\delta^{13}\text{C} \sim -18$ to -21) (ref. 12). The $\delta^{13}\text{C}$ data of sediments in this study range from -17.8 to 21.8% , indicating a marine source. This is consistent with lignin studies which have shown that land-derived organic matter is typically not an important component of surface marine sediments at distances > 50 -100 km from shore^{13,14}, and would appear to rule out a major terrestrial component for the organic carbon in these sediments.

The ^{14}C activity of the sedimentary organic matter is a powerful tool for distinguishing the remaining sources because their ^{14}C ages span the entire range from modern to undetectable. Plankton records the radiocarbon activity of marine surface waters which was 95% modern until the onset of atmospheric nuclear-weapons testing in the mid-1950s and has since risen to 109-121% modern¹⁵; DOC in the ocean has a mean activity of 66% modern¹⁶; and fossil organic carbon from sedimentary rocks is effectively 'dead'. Limits can be placed on the importance of the first two of these sources by using knowledge of their chemical reactivity and a mass balance model of carbon flow in surface sediments. The low organic-carbon content of continental sedimentary rocks coupled with the very low clay content in CaCO_3 -rich sediments and the advanced mixed layer age of CaCO_3 -poor sediments prevents us from distinguishing the presence of fossil organic carbon from the ^{14}C data.

The mean ages of the surface sediments in the north-west Atlantic and equatorial Pacific (Fig. 1) are younger than the mean deep-ocean DOC age (3,400 yr BP), indicating the necessity for a recent component in the sediments. Also, the first-order correlation of organic carbon age with sediment residence time in the mixed layer (Fig. 1) implies that there is a carbon fraction within the sediments with a degradation lifetime greater than the mixed layer residence time. Organic carbon from DOC sources should occupy the more refractory com-

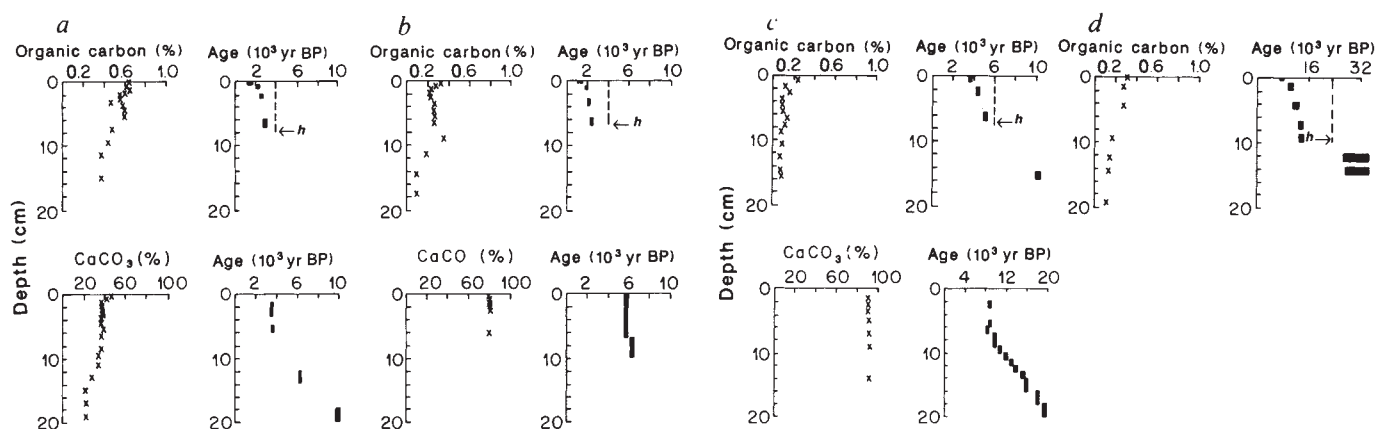


Fig. 1 Depth profiles of organic carbon concentration, organic ^{14}C age, CaCO_3 content and CaCO_3 ^{14}C age for four locations in the ocean: *a*, north-west Atlantic (41°N , 64°W , 3,000 m); *b*, equatorial Pacific (1°N , 134°W , 4,400 m); *c*, South Pacific (12.3°S , 125.9°W , 3,888 m); and *d*, North Pacific. (A data table is available on request from the authors.) Arrows indicate the mixed-layer depth, h , determined from the CaCO_3 ^{14}C (*a-c*) or thorium-230 (*d*) activities. The dashed vertical line is the predicted refractory organic ^{14}C age from the equation $t_R = 1/\lambda \ln [s/(s + \lambda h)]$, where s is the sedimentation rate and $\lambda = 1.24 \times 10^{-4} \text{ yr}^{-1}$. Sedimentation rates for *a-d*, respectively are: 2.0, 1.4, 0.9 and 0.08 cm per 1,000 yr. These values were determined from the CaCO_3 ^{14}C stratigraphy or from published values. The data from the North Pacific are from three cores, reported by Druffel *et al.*². CaCO_3 ^{14}C values from the South Pacific were determined by S. Langer and W. S. Broecker (unpublished results) and from the equatorial Atlantic by Berger and Killingley²⁴.

Methods. Sediment cores were stored frozen and in the dark after retrieval and sectioning. Samples for AMS ^{14}C analysis were thawed, dried, acidified and combusted at 800°C in a closed quartz tube in the presence of MnO_2 and CuO (ref. 8). CO_2 was reduced to graphite by iron-catalysed hydrogen reduction and analysed for ^{14}C content on the Van de Graaff accelerator in the University of Washington Nuclear Physics Laboratory^{9,10}. The ^{14}C ion beam of the sample is compared with that of a working standard, each normalized to its respective ^{13}C ion beam. The ^{14}C concentration ratio is converted to the conventional radiocarbon ratio¹¹, $A_{\text{SN}}/A_{\text{ON}}$, by correcting for (1) isotope fractionation to $\delta^{13}\text{C} = -25\%$ for samples and $\delta^{13}\text{C} = -19\%$ for the standard, (2) machine background and (3) sample preparation contamination determined from two identical preparations of anthracite. A_{SN} in the conventional radiocarbon ratio is the normalized sample ^{14}C activity and A_{ON} is $0.95 \times$ the normalized NBS oxalic acid CO_2 ^{14}C activity. Two anthracite results ($A_{\text{SN}}/A_{\text{ON}} = 0.0189 \pm 0.0010$ apparent age $31,890_{-422}^{+450}$ yr, and $A_{\text{SN}}/A_{\text{ON}} = 0.0218 \pm 0.0015$, apparent age $30,750_{-530}^{+570}$ yr) indicated contamination during sample preparation equivalent to the admixture of 1.9% of carbon of activity A_{ON} . The conventional radiocarbon age is $t = -(1/\lambda) \ln (A_{\text{SN}}/A_{\text{ON}})$ yr BP, where $\lambda = 1.24 \times 10^{-4} \text{ yr}^{-1}$ and $A_{\text{SN}}/A_{\text{ON}} \sim \%$ modern. The ^{14}C uncertainty intervals in the figures correspond to standard deviations based either on the variance in the distribution of repeated measurements of the same sample material, or on the uncertainty of individual measurements, whichever is larger. In the latter case, s.d. is determined from results of 15 to 20 half-minute data collection cycles that make up a single measurement.

ponent of the sediment organic-carbon reservoir (that is, that which is buried below the mixed layer) because this source is relatively inert to degradation on the timescale of the mixed layer residence time in all but the North Pacific cores.

Because the organic-carbon profiles and the organic- ^{14}C data reveal characteristics of both a refractory (R), and a degradable (D) fraction, we assume a working model in which the mean concentration, $\bar{C}$, of particulate organic carbon in the sediment mixed layer is a two-component mixture:

$$\bar{C} = \bar{C}_R + \bar{C}_D \quad (1)$$

D indicates 'mixed layer degradable', and, by our definition, means that the lifetime with respect to *in situ* degradation is of the same magnitude or less than the residence time in the mixed layer. In reality there is probably a continuum of degradation

rates¹⁷ and the value assigned to $\bar{C}_D$ is a combination of these. However, the difference between 'true' and mean reservoir size caused by using a mean degradation lifetime¹⁸ is greatly limited here because the maximum age for any component of $\bar{C}_D$ is the mixed layer residence time, τ_p .

In the depth interval that is mixed by benthic animals we describe the steady state mass balance for the average mixed layer organic carbon concentration, $\bar{C}$, as the equality between the flux of sinking particulate organic carbon to the sediment surface, F , and the removal rates of carbon by sedimentation, s , and reaction, γ .

$$F = s\bar{C} + \gamma h\bar{C} \quad (2)$$

Four equations of this form can be written to describe the average concentrations of mixed layer degradable and refractory carbon

Table 1 Calculation of the mixed-layer degradable organic carbon in the world's oceans

Ocean province	Area (km^2)*	Average organic-carbon content (% d.w.)†	Mixed-layer depth (cm)	Degradable fraction in mixed-layer total	Mixed-layer total (g)
Continental margin	42×10^6	1.02	20‡	0.3–0.5	$19\text{--}32 \times 10^{15}$
Abyssal ocean	320×10^6	0.34	10	0.2–0.4	$17\text{--}33 \times 10^{15}$
				Total	$50 \pm 15 \times 10^{15}$

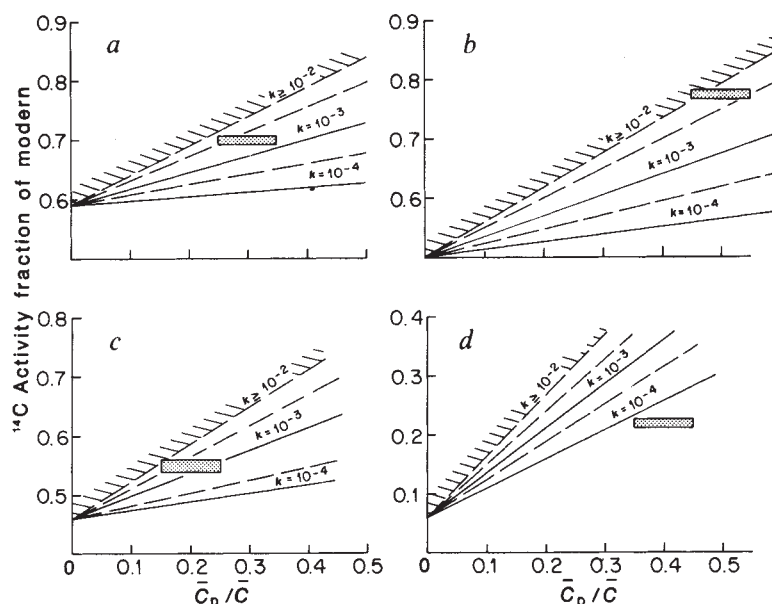
In this compilation we have included the slow sedimentation areas of the abyssal ocean, which may be unrealistic for our purpose because of the long degradation residence time (Fig. 2); however, it is a small component of the total because of the low content. We have also deleted from consideration organic carbon in the sediments of the adjacent seas of the ocean, which by some estimates¹⁹ could constitute up to one third of the global marine sedimentary carbon reservoir. We estimate that the mixed-layer depth and fraction degradable to be lower limits for the continental margins.

* The continental margin consists of the continental shelf, slope and rise, Seibold and Berger²⁵.

† Romenkevich¹⁹ calculated the average organic-carbon content in the Atlantic and Pacific Oceans. We assume that it is the same in the Indian Ocean.

‡ Based on the maximum depth of penetration of excess ^{210}Pb and Pu (ref. 3).

Fig. 2 The mean mixed layer organic carbon ^{14}C activity versus the fraction of the organic carbon reservoir that is degradable, C_D/C , at four locations in the ocean: *a*, north-west Atlantic; *b*, equatorial Pacific; *c*, South Pacific; *d*, North Pacific. Rectangles represent the data. For *a-d*, the mean organic ^{14}C activities are: 0.71 ± 0.015 , 0.78 ± 0.015 , 0.56 ± 0.15 and 0.22 ± 0.02 (fraction modern) respectively, and the ratios C_D/C are: 0.3, 0.5, 0.2 and 0.4 with an estimated uncertainty of ± 0.05 . Lines for different degradation rate constants k (yr^{-1}) were derived from equation (3) with mixed-layer depths, h , and sedimentation rates, s , given in Fig. 1. Solid lines are for a degradable organic-carbon source that is equation prebomb ($C_D^* = 0.95$ modern); dashed lines are for a postbomb source ($C_D^* = 1.10$ modern). Refractory organic carbon, $\bar{C}_R$, is assumed to have a modern, prebomb activity for these solutions. An activity consistent with the age of DOC (3,400 yr BP, $C_R^* = 0.66$ modern) would lower the lines by 0.09–0.18 modern in *a*; 0.07–0.11 in *b*; 0.09–0.14 in *c*; and 0.01–0.02 in *d*. With the exception of *d*, a DOC age for $\bar{C}_R$ could not explain the measurements. The North Pacific data is insensitive to the initial age of $\bar{C}_R$ because of the low sedimentation rates.



($\bar{C}_D$, $\bar{C}_R$) as well as radioactive and stable carbon (^{14}C , C). The corresponding reaction terms, γ , consist of different combinations of the organic matter degradation constant, k , and radioactive decay constant, λ (for example, $\bar{C}_R$, $\gamma = 0$; $^{14}\bar{C}_R$, $\gamma = \lambda$; C_D , $\gamma = k$; $^{14}C_D$, $\gamma = k + \lambda$). Combination of the four equations and (1) allows us to write an expression for the $^{14}\bar{C}/^{12}\bar{C}$ ratio of the mixed layer, C^* , in terms of the isotope ratio of the particulate flux, C^*F :

$$C^* = (\bar{C}_D/\bar{C})C_D^*F/(1 + \lambda h/(s + kh)) + (\bar{C}_R/\bar{C})C_R^*F/(1 + \lambda h/s) \quad (3)$$

The relation between C^* and the fraction of the mixed-layer organic carbon that is degradable, $\bar{C}_D/\bar{C}$, is plotted in Fig. 2; for different organic-matter degradation constants, k . The $^{14}\text{C}/^{12}\text{C}$ ratio of the particulate flux in this plot is assumed to be either prebomb or postbomb for the degradable component. The refractory component is assigned a modern, prebomb activity for this calculation; the ramifications of greater age corresponding to DOC is discussed below and in the caption. Measured mixed layer average ^{14}C activities and the C_D/C ratios are indicated as rectangles in the figures. In three out of the four cases (the exception is the North Pacific core) the organic ^{14}C activities are near the maximum obtainable by this model. The results indicate residence times ($1/k$) for the degradable organic carbon fraction of $<1,000$ yr except in the North Pacific core where it is greater than 10,000 yr. A greater age for the source of the refractory organic carbon would lower the model-derived lines as indicated in the Fig. 2 legend. The ^{14}C ages in three of the four locations are too young to accommodate a significant organic carbon fraction with a seawater DOC source.

The rate of degradation is sufficiently fast that the degradable organic carbon concentration of marine sediments should have responded to changing oceanic conditions from glacial to interglacial periods. The global size of this reservoir is estimated in Table 1. Average surface-sediment organic carbon concentrations for the continental margin and abyssal ocean were taken from recent compilations^{19,20}. Because none of our measurements are from the continental shelf or slope, we assume a mixed-layer depth based on the published distribution of short-lived radioisotopes³ and a degradable organic carbon component similar to the north-west Atlantic and equatorial Pacific locations. Both of these estimates are believed to be lower limits. Notice that half of the mixed layer degradable organic carbon is presently stored in the continental margins even though it is

only about one eighth of the ocean bottom area. We estimate a present reservoir size, $\Sigma \bar{C}_D$, of $50 \pm 15 \times 10^{15}$ g of mixed layer degradable carbon.

Although there is insufficient data from sediment cores to determine the history of global change in the sedimentary organic carbon reservoir, and the magnitude of the signal may have been eroded by degradation²¹, measurements^{22,23} indicate that the concentrations were as much as four to five times higher in continental margin and upwelling areas during the past glacial period. If this is worldwide, then the sedimentary organic carbon reservoir could have been $150\text{--}200 \times 10^{15}$ g C greater ($\Delta \Sigma \bar{C}_D$) in glacial times, which would be equivalent to a change of roughly $15 \mu\text{M}$ AOU if oxidized in the deep ocean.

We acknowledge financial support from the University of Washington College of Ocean and Fishery Sciences NSF and the US Department of Energy. Lewis Lo carried out careful organic carbon analyses and David Balsley and Donald Leach prepared the targets for AMS analyses. The helpful comments of J.I. Hedges are greatly appreciated.

Received 9 March; accepted 12 June 1987.

- Williams, P. M., Stenhouse, M. C., Druffel, E. M. & Koide, M. *Nature* **276**, 698–701 (1978).
- Druffel, E. M., Williams, P. M., Livingston, H. D. & Koide, M. *Earth planet. Sci. Lett.* **71**, 205–214 (1984).
- Walsh, J. J. *et al. Deep Sea Res.* **32**, 855–993 (1985).
- Tissot, B. B. & Welte, D. H. *Petroleum Formation and Occurrence* (Springer, Berlin, 1978).
- Hunt, J. M. *Bull. Am. Ass. Petrol. Geol.* **56**, 2273–2277 (1972).
- Emerson, S. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. & Broecker, W. S.) (AGU Monograph 32, Washington DC, 1985).
- Smith, K. L. & Baldwin, R. J. *Nature* **307**, 624–626 (1984).
- Buchanan, D. L. & Corcoran, B. J. *Analyt. Chem.* **31**, 1635–1638 (1959).
- Farwell, G. W., Grootes, P. M., Leach, D. D., Schmidt, F. H. & Stuiver, M. *Radiocarbon* **25**, 711 (1983).
- Farwell, G. W., Grootes, P. M., Leach, D. D. & Schmidt, F. H. *Nucl. Instrum. Meth.* **233** (B5), 144 (1984).
- Stuiver, M. & Polach, M. A. *Radiocarbon* **19**, 255 (1977).
- Degens, E. T. in *Organic Geochemistry* (eds Eglinton, G. & Murphy, E. T. J.) 304–329 (Springer, Berlin, 1969).
- Hedges, J. T. & Parker, P. L. *Geochim. cosmochim. Acta* **43**, 1019–1029 (1976).
- Hedges, J. T. & Mann, D. C. *Geochim. cosmochim. Acta* **43**, 1803–1809 (1979).
- Druffel, E. M. & Linick, T. W. *Geophys. Res. Lett.* **5**, 913–916 (1978).
- Williams, P. M., Oechger, H. & Kinney, P. J. *Nature* **224**, 256–258.
- Jorgensen, B. B. *Geomicrobiol. J.* **1**, 29–47 (1979).
- Mantoura, R. F. & Woodward, E. M. S. *Geochim. cosmochim. Acta* **47**, 1293–1309 (1983).
- Romankevich, E. A. *Geochemistry of Organic Matter in the Ocean* (Springer, Berlin, 1984).
- Premuzic, E. T., Benkovitz, C. M., Guftney, J. S. & Walsh, J. J. *Org. Geochem.* **4**, 63–77 (1982).
- Wilson *et al. Science* **232**, 972–975 (1986).
- Muller, P. J. & Suess, E. *Deep Sea Res.* **26**, 1347–1367 (1979).
- Petersen, T. F. *Geology* **11**, 16–19 (1983).
- Berger, W. & Killingley, J. S. *Mar. Geol.* **45**, 93–125 (1982).
- Seibold, E. & Berger, W. H. *The Sea Floor* (Springer, Berlin, 1982).

Occurrence and geochemical significance of long-chain dialkylthiacyclopentanes

J. C. Schmid*, J. Connan† & P. Albrecht*

* Département de Chimie, Université Louis Pasteur, 1, rue Blaise Pascal 67008 Strasbourg, France

† Centre de Recherches, Société nationale Elf-Aquitaine (Production) Chemin Micouleau, 64000 Pau, France

Small sulphur-containing molecules were first detected in petroleum and sediments many years ago¹⁻³, but only recently have complex sulphur components related to biological markers been identified⁴⁻¹⁴. These discoveries have provided interesting information on the distribution of sulphur-containing compounds in geological samples, but the formation mechanisms of such molecules remain to be clarified. Recently, Connan *et al.*¹⁵ recognized the occurrence of series of long-chain sulphur-containing components in heavy sulphur-rich crude oils from carbonate basins and tentatively identified them as alkylthiacyclopentanes and alkylthiacyclohexanes on the basis of selective detection and gas chromatography/mass spectrometry studies. Here we report the conclusive identification and geochemical significance of long-chain 2,5-dialkylthiacyclopentanes of type 1 (see Fig. 4), occurring as major components in some of these petroleum, as shown for an immature carbonate crude oil from Maruejols (Oligocene Alès basin, France), containing 6.5% sulphur¹⁶. In addition to providing new information about the 'aromatic' fractions of these crude oils (which are in fact largely non-aromatic), these results give useful indications of some of the mechanisms of incorporation of sulphur into organic matter in the sub-surface environment. Furthermore, as confirmed by thermal simulation experiments, these dialkylthiacyclopentanes are thermally labile, which seems to restrict their occurrence to immature crude oils. They could therefore serve as useful maturation parameters for the early stages of petroleum formation in carbonate basins.

The crude oil of Maruejols (1.15 g) was first crudely separated on a silica-gel column (Merck) and a non-polar fraction collected by elution with hexane, then hexane-ether (99:1). This fraction was rechromatographed on another silica-gel column. Elution with hexane gave mainly saturated hydrocarbons;

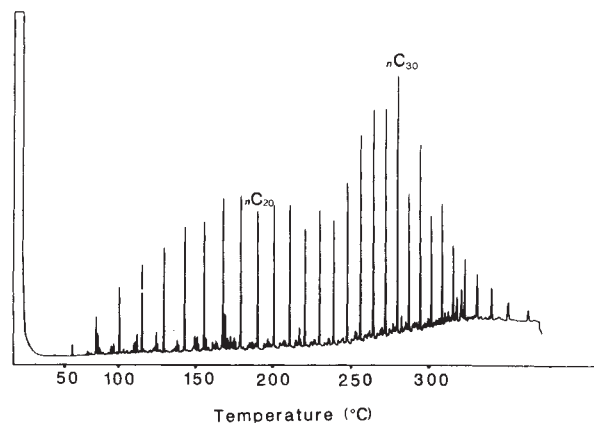


Fig. 2 Gas chromatogram of total alkanes obtained from desulphurization of total aliphatic sulphides of Maruejols crude oil with Raney nickel. Conditions: OV 101, 40 m × 0.32 mm, 40–300 °C, 4 K min⁻¹.

further elution with hexane, followed by hexane-ether (99:1) yielded 440 mg of a fraction containing aliphatic and aromatic sulphides, as well as some aromatic hydrocarbons. This mixture was further separated by thin-layer chromatography (TLC) on silica gel impregnated with 1% silver nitrate with methylene chloride as an eluting solvent. This method gave a clear separation, by increasing polarity, of total aromatics (sulphides and hydrocarbons) and aliphatic sulphides. The latter represented about 9.5% of the total crude oil.

The gas chromatogram of the aliphatic sulphide fraction of Maruejols crude oil (Fig. 1) showed a succession of regular patterns between C₁₀ and C₄₀, which suggested homologous series. For the major series gas chromatography/mass spectrometry showed the occurrence of fragments typical of thiacyclopentanes¹⁷, presumably polyalkylated (C₄H₇S⁺, *m/z* = 87; *m/z* = 87 + 14*n*), and revealed the presence of a great number of isomers in each pattern. Reductive desulphurization of this extremely complex mixture with Raney nickel (Fluka), however, gave a surprisingly simple saturated hydrocarbon fraction, in quantitative yield, which was largely dominated by *n*-alkanes and closely resembled the alkane fraction of the oil (Fig. 2). The simplification of the distribution pattern suggested that the major components of this mixture were thiacyclopentanes of type 1, monoalkylated in position 2 or dialkylated in positions 2 and 5. Thus, the large number of isomers might be explained by a random position for sulphur in the alkyl chains (for example for the C₂₀ pattern of 1 all possible dialkyl isomers compatible with *m* + *n* = 14 would be present). The reduction of such a mixture with Raney nickel would indeed essentially yield *n*-alkanes. This feature was not only the key to the structural elucidation of these compounds, but was also indicative of a possible manner of formation¹⁶. Other authors have made a similar observation on several crude oils, although limited to shorter chain compounds, which were tentatively identified as α -alkyl- or α , α' -dialkylthiamonocyclanes on the basis of formation of *n*-alkanes in desulphurization reactions¹.

To confirm this hypothesis we synthesized eight 2-alkylthiacyclopentanes and 2,5-dialkylthiacyclopentanes between C₁₈ and C₂₃ as reference compounds (Fig. 3), in using 2-formyl-thiophene, 2-formyl-5-methyl-thiophene and 2-bromo-5-formyl-thiophene (Fluka)¹². All reactions resulted in fairly good yields, and the final and intermediate products displayed expected spectral and analytical data. Reduction of intermediate thienyl alcohols was done following an ionic hydrogenation procedure developed by Kursanov *et al.*¹⁸ Depending on the reaction conditions and the reaction times, this yields either the pure thiophenes or a mixture of thiophenes and thiacyclopent-

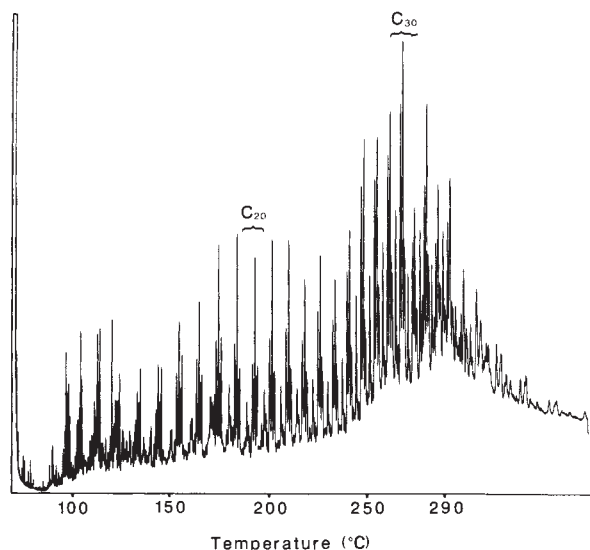


Fig. 1 Gas chromatogram of total aliphatic sulphides of Maruejols crude oil. Conditions: SE 54, 37 m × 0.32 mm; 40–100 °C, 5 K min⁻¹ and 100–290 °C, 2.5 K min⁻¹.

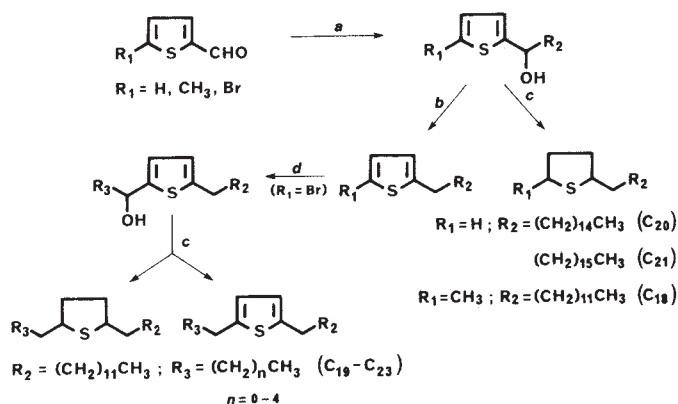


Fig. 3 Synthetic scheme for the preparation of 2-alkyl- and 2,5-dialkylthiacyclopentanes and the corresponding thiophenes. *a*, R₂MgBr, Et₂O; *b*, CF₃CO₂H, HSiEt₃, CH₂Cl₂, room temperature 5 min; *c*, CF₃CO₂H, HSiEt₃, 50 °C, 48 h; *d*, BuLi, -78 °C; R₃CHO. Dialkylthiacyclopentanes are obtained as *cis* + *trans* mixtures.

tanes (about 1:1 after 24 h) easily separable by TLC on silica gel. All 2,5-dialkylthiacyclopentanes were obtained as a mixture of *cis* and *trans* isomers, which were resolved by capillary column GC when one of the alkyl substituents had > two carbons. Both isomers displayed identical mass spectra. We could however, tentatively attribute the *cis* configuration to the isomers that eluted first on a non-polar column because they are more bulky and should therefore interact less with the phase. This was confirmed by varying the hydrogenation procedure of the C₂₂ intermediate thiophene (2-tridecyl-5-pentyl-thiacyclopentane). Although the ionic reduction gave a 35/65 mixture of *cis* and *trans* thiacyclopentane, a catalytic reduction with palladium on charcoal yielded, as expected, a predominance of the first eluting *cis* isomer (*cis*/*trans* = 85/15).

Mass spectral data of the synthetic compounds were as follows (EI, 70 eV). 2-hexadecylthiacyclopentane, *m/z* = 312 (M⁺, 12%), 269(12), 255(2), 227(5), 213(2), 185(3), 171(5), 143(7), 129(20), 101(10), 87(100); 2-heptadecylthiacyclopentane, 326 (M⁺, 5), 283(4), 241(2), 171(2), 129(9), 101(5), 87(100); 2-methyl-5-tridecylthiacyclopentane, 284 (M⁺, 12), 227(14), 185(8), 157(4), 143(16), 101(100), 87(5); 2-ethyl-5-tridecylthiacyclopentane, 298 (M⁺, 12), 269(29), 227(13), 199(3), 185(4), 171(2), 157(11), 115(100), 87(17); 2-propyl-5-tridecylthiacyclopentane, 312 (M⁺, 19), 269(51), 227(18), 213(5), 185(10), 129(100), 101(6), 87(41); 2-butyl-5-tridecylthiacyclopentane, 326 (M⁺, 23), 269(67), 227(25), 185(16), 143(100), 101(26), 87(53); 2-pentyl-5-tridecylthiacyclopentane, 340 (M⁺, 23), 311(8), 269(82), 241(6), 227(23), 213(7), 199(19), 185(7), 157(100), 115(25), 87(72); 2-hexyl-5-tridecylthiacyclopentane, 354 (M⁺, 25), 311(13), 269(90), 255(7), 227(26), 213(20), 185(7), 171(100), 129(25), 87(71); 2-methyl-5-tridecylthiophene: 280 (M⁺, 22), 111(100).

The alkyl- and dialkylthiacyclopentanes displayed very peculiar fragmentation patterns leading to series of fragments separated by 42 AMU. Indeed two series of fragments can generally be observed, for example for 2-hexadecylthiacyclopentane: (i) 87 + *n*42 (ii) M⁺ - 43 - *n*42. A mechanism implying the migration of a hydrogen from position 3' to 5 could explain their occurrence and could eventually be checked on deuterated compounds.

All the 2,5-dialkylthiacyclopentanes synthesized coeluted with major peaks in the gas chromatogram of the sulphide mixture of Maruejols crude oil. The various isomers of C₁₈ and C₁₉ compounds coeluted as single peaks (*cis* and *trans* isomers not resolved), but there is excellent agreement between their mass spectra and those of the geochemical mixture. The other dialky-

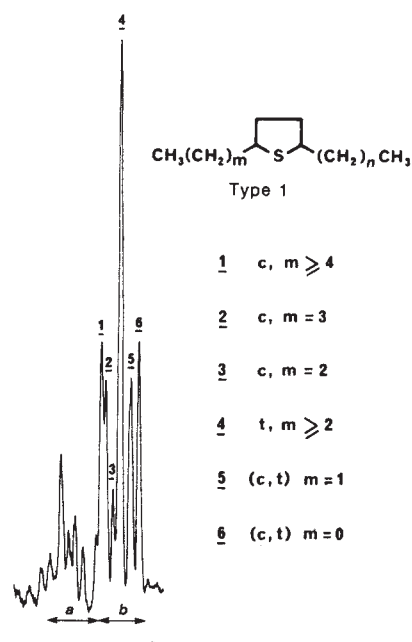


Fig. 4 Partial extended gas chromatogram of total aliphatic sulphides of Maruejols crude oil showing the distribution of C₂₀ dialkylthiacyclopentanes. Conditions: as in Fig. 1. Peaks: *a*, thiacyclohexanes; *b*, thiacyclopentanes; *m* + *n* = 14, *n* > *m*. The following are not separated: (i) *trans* isomers for *m* > 2; (ii) *cis* isomers for *m* > 4; (iii) *cis* from *trans* isomers for *m* = 0 and *m* = 1.

lated compounds (C₂₀-C₂₃) synthesized coeluted as doublets of *cis* and *trans* isomers, but it was not possible to correlate directly their mass spectra with those of the sulphides of the crude oil because all the isomers of the latter could not be resolved on the capillary columns used. The peaks corresponding to mixtures usually show fragmentation patterns corresponding to several isomers, which yield mass spectra displaying fragments every 14 AMU (*m/z* = 87 + 14*n*). The 2-alkylthiacyclopentanes do not coelute with any major peaks in the geochemical mixture; they can only be present, if at all, in trace amounts. Figure 4 summarizes our observations based on GC retention times and mass spectrometry; it is obvious that the *trans* isomers are dominant. (When this work was essentially completed we became aware that the Delft group had tentatively identified 2,5-dialkylthiacyclopentanes in Rozel Point crude oil; these authors have now further confirmed their observations by synthesis^{11,19}.)

The smaller repetitive patterns appearing between the homologous dialkylthiacyclopentanes display mass spectral fragmentation patterns corresponding to dialkylthiacyclohexanes. Our results for the thiacyclopentanes and from Raney nickel reduction suggest that they are 2,6-dialkyl thiacyclohexanes, a hypothesis which needs confirmation by synthesis (see also ref. 19).

Complex functionalized molecules can incorporate sulphur in sediments at early stages of diagenesis, presumably by reaction with hydrogen sulphide (or related species) or with elemental sulphur, both abundantly produced by bacterial processes in anoxic environments^{4,6,10-14} and available for combination with organic matter, especially in carbonates where little iron is usually present for trapping these active sulphur species. Hence, diagenetic and maturation processes could, from common biological precursors, yield alkanes on one side and thiacyclopentanes on the other. However, the similarity between the alkanes of the Maruejols crude oil and those obtained by Raney nickel reduction of its sulphides, along with the random incorporation of sulphur into the alkane chains suggest a possible alternative formation of the alkylated thiacyclopentanes by a

reaction between the alkanes and elemental sulphur¹² presumably during early maturation in the source rock (the oil indeed shows distinctive signs of immaturity), rather than between sulphur and functionalized molecules. To confirm this hypothesis we did a simulation experiment in which we heated *n*-octadecane in the presence of sulphur at temperatures varying between 200 °C and 250 °C in a sealed glass tube for periods of ~65 h. An aromatic sulphide fraction (6–10% of starting material) was formed, which comprised dialkylthiophenes as well as alkylated polythienyls. Although the intermediate alkylated thiacyclopentanes could not be trapped under the reaction conditions used, the monothiophene fraction did consist almost exclusively of a random mixture of C₁₈ 2,5-dialkylthiophenes, in support of the above hypothesis¹². Small amounts of thiacyclopentanes were formed, however, by heating pristane and 1-hexadecene under similar conditions¹². Furthermore, the small amounts of dialkylbenzothiophenes formed in this reaction triggered our interest in this class of compounds²⁰. Their identification, their distribution in geological samples, as well as their presumable formation by further cyclization/aromatization of 2,5-dialkylthiophenes will be described elsewhere (N. Perakis, J. C. S. and P. A. in preparation; see also ref. 19).

A similar relation between alkanes and thiacyclopentanes was established for Rozel Point seep oil, a sulphur-rich highly immature petroleum from Utah, USA (15% S), although in this case the aliphatic sulphides were largely dominated by polycyclic sulphur containing biological markers, in particular thiosteroids¹².

The same 2,5-dialkylthiacyclopentane series were further observed in several other carbonate crude oils. However, they occurred in significant amounts only in quite immature petroleum (such as the Gallician, Tertiary of Camargue basin, France; Amposta, Spain; Gela, Sicily). They seem therefore to be transient species which are transformed under increasing thermal stress into dialkylthiophenes or other more aromatic organo-sulphur compounds which are found in more mature crude oils. Their rapid disappearance with increasing maturation was further confirmed by thermal simulation experiments carried out on Rozel Point¹² and Maruejols crude oils. Hence, this class of compounds shows great potential as molecular maturation parameters for the early phase of petroleum formation, especially in carbonate basins, a point which is now submitted to further investigations.

We thank Dr J. de Leeuw for helpful discussions and exchange of information, Dr G. Teller and P. Wehrung for mass spectral analysis and Mrs E. Krempf for NMR measurements.

Received 3 February; accepted 5 June 1987.

1. Aksenov, V. S. & Kamyarov, V. F. in *Organic Sulfur Chemistry* (eds Freidlina, R. Kh. & Skorova, A. E.) 1–13 (Pergamon, Oxford, 1981).
2. Thompson, C. J. in *Organic Sulfur Chemistry* (eds Freidlina, R. Kh. & Skorova, A. E.) 189–208 (Pergamon, Oxford, 1981).
3. Whelan, J. K., Hunt, J. M. & Berman, J. *Geochim. cosmochim. Acta* **44**, 1767–1785 (1985).
4. Rullkotter, J., Von der Dick, H. & Welte, D. H. in *Init. Rep. DSDP Vol. 63* (eds Haq, B. et al.) 819–836 (US Government Printing Office, Washington DC, 1981).
5. Payzant, J. D., Montgomery, D. S. & Strausz, O. P. *Tetrahedron Lett.* **24**, 651–654 (1983).
6. Valisollalao, J., Perakis, N., Chappe B. & Albrecht, P. *Tetrahedron Lett.* **25**, 1183–1186 (1984).
7. Payzant, J. D., Cyr, T. D., Montgomery, D. S. & Strausz, O. P. *Tetrahedron Lett.* **26**, 4175–4178 (1985).
8. Cyr, T. D., Payzant, J. D., Montgomery, D. S. & Strausz, O. P. *Org. Geochem.* **9**, 139–143 (1986).
9. Payzant, J. D., Montgomery, D. S. & Strausz, O. P. *Org. Geochem.* **9**, 357–369 (1986).
10. Brassel, S. C., Lewis, C. A., de Leeuw, J. W., de Lange F. & Sinninghe Damste, J. S. *Nature* **320**, 160–162 (1986).
11. Sinninghe Damste, J. S., ten Haven, H. L., de Leeuw, J. W. & Schenck, P. A. in *Advances in Organic Geochemistry 1985* (eds Leythaeuser, D. & Rullkotter, J.) *Org. Geochem.* **10**, 791–805 (1986).
12. Schmid, J. C. thesis, Univ. Louis Pasteur, Strasbourg (1986).
13. Sinninghe Damste, J. S. & de Leeuw, J. W. *Int. J. Environ. anal. Chem.* (in the press).
14. Sinninghe Damste, J. S., Kock-van Dalen, A. C., de Leeuw, J. W. & Schenck, P. A. *Tetrahedron Lett.* (in the press).
15. Connan, J., Grondin, J. L., Colin, J. P., Hussler, G. & Albrecht, P. *Communication at 11th Int. Meet. Org. Geochem.* (The Hague, September, 1983).
16. Ocampo, R. et al. *Communication at 12th Int. Meet. Org. Geochem.* Jülich, September 1985.
17. Budzikiewicz, H., Djerassi, C. & Williams, D. H. *Interpretation of Mass Spectra of Organic Compounds* 284–286 (Holden-Day, San Francisco 1964).
18. Kursanov, D. N., Parnes, Z. N. & Loim, N. M. *Synthesis*, 633–650 (1974).
19. Sinninghe Damste, J. S. et al. *Geochim. cosmochim. Acta.* (in the press).
20. Perakis, N. thesis, Univ. Louis Pasteur, Strasbourg (1986).

Rapid removal of Chernobyl fallout from Mediterranean surface waters by biological activity

Scott W. Fowler*, Patrick Buat-Menard†, Yuji Yokoyama†, Serge Ballestra*, Elis Holm* & Huu Van Nguyen†

* International Laboratory of Marine Radioactivity, International Atomic Energy Agency, c/o Musée Océanographique, MC 98000 Monaco

† Centre des Faibles Radioactivités, Laboratoire Mixte CNRS-CEA, BP 1, F-91190, Gif-sur-Yvette, France

The sinking of particulate organic matter from the euphotic zone is an important pathway for the vertical transport of many elements and organic compounds in the sea^{1–3}. Many natural^{4–5} and artificial^{5–7} radionuclides in surface waters are readily adsorbed onto suspended particles and are presumably scavenged and removed to depth on time scales commensurate with both particle sinking rate and retention time of the radionuclide on the particle. Previously, abyssal benthic organisms from the northeast Pacific were found to contain short-lived fission products which entered the sea surface as fallout from nuclear testing⁸. The presence of these radionuclides at great depth could not be explained by Stokesian settling of small fallout particles and it was hypothesized⁸ that zooplankton grazing in the surface layers packaged these particle-reactive radionuclides into large, relatively dense faecal pellets which rapidly sank to depth. We report here data from a time-series sediment trap experiment and concomitant zooplankton collections which show conclusively that Chernobyl radioactivity, in particular the rare earth nuclides ¹⁴¹Ce and ¹⁴⁴Ce, entering the Mediterranean as a single pulse, was rapidly removed from surface waters and transported to 200 m in a few days primarily by zooplankton grazing.

Experiments to measure the vertical flux of selected elements and compounds introduced into surface waters from the atmosphere were carried out in the northwestern Mediterranean between 13 April and 21 May 1986. An automated time-series particle trap⁹ (A. Monaco et al., in preparation) was moored at a depth of 200 m in a 2,200-m water column ~15 nautical miles off the coast of Calvi, Corsica (42°43.9' N 08°31.3' E). Each of the six trap collector cups contained a buffered formalin solution to avoid decomposition losses¹⁰, and was set to sample sinking particles for consecutive periods of 6.25 days. On 26 April 1986 the nuclear accident at Chernobyl occurred which resulted in relatively high levels of radioactive fallout over much of Europe^{11–13} including the northwestern Mediterranean¹⁴. Air-filter measurements¹⁴ (N. E. Whitehead et al., in preparation) made at Monaco, 77 nautical miles north-west of the particle trap site, first detected increased atmospheric radioactivity on 30 April with a maximum around 3 May. Furthermore, rain over the next two days accounted for 80% of the total radionuclide deposition at Monaco¹⁴. One week later, air concentrations had dropped to 1% of the peak values indicating that fallout entered the sea in this region essentially as a single pulse during the period 4–5 May.

The particle trap was retrieved on 22 May and the particulate material, along with other marine samples, was immediately returned to shore and analysed for radioactivity by non-destructive γ -spectrometry (Table 1). The results indicated that some radioactivity reached 200 m very soon after the peak fallout entered the surface waters (see sample 4). However, the primary pulse of particulate radionuclides arrived at that depth between 8 and 15 May, on the average seven days after the peak radioactivity was delivered to the surface. This time lag implies a mean sinking speed for the radioactive particles of roughly 29 m d⁻¹. The pulsed input to 200 m was particularly evident

Table 1 Prominent Chernobyl fallout radionuclides in large particles collected at 200 m

Sample:	1	2	3	4	5	6
Date:	13-20 April	20-26 April	26 April-2 May	2-8 May	8-15 May	15-21 May
Dry weight (g):	0.1678	0.0875	0.05017	0.05143	0.04206	0.04526
Radionuclide						
⁹⁵ Zr	<0.07	<0.2	<0.3	<0.2	24.5 ± 1.4	<0.2
⁹⁵ Nb	<0.03	<0.1	<0.2	<0.1	31.8 ± 1.1	<0.2
¹⁰³ Ru	<0.06	<0.1	<0.2	3.7 ± 0.2	23.6 ± 1.0	14.0 ± 0.4
¹⁰⁶ Ru	<0.2	<0.4	<0.8	1.1 ± 0.5	5.4 ± 1.8	3.5 ± 0.7
¹³⁴ Cs	<0.05	<0.05	<0.05	0.41 ± 0.05	2.1 ± 0.2	1.9 ± 0.1
¹³⁷ Cs	<0.05	<0.05	0.15 ± 0.08	0.85 ± 0.08	3.8 ± 0.3	4.0 ± 0.1
¹⁴¹ Ce	<0.2	<0.3	<0.3	1.3 ± 0.7	12.6 ± 0.6	1.1 ± 0.5
¹⁴⁴ Ce	<0.06	<0.3	<0.3	<0.2	13.6 ± 0.7	<0.4

Values are as Bq per gram dry weight collected in a particle interceptor trap moored in water 2,200 m deep off the northwest coast of Corsica in April and May 1986. Sinking particulates were collected in a cylindrical sediment trap with an 0.125-m² opening and height-diameter aspect ratio of 2.5. The trap was beneath the base of the euphotic zone which in this region normally lies between 80 and 100 m deep²⁸. Samples were analysed by γ -spectrometry at both ILMR and Centre des Faibles Radioactivités (CFR) laboratories using Ge (Li) or high purity (HP) Ge detectors^{14,29}. Sample radioactivity was calibrated by standard reference sources (Amersham) of known absolute emission. In most cases agreement was good between the two laboratories. Most of the values in the table are from CFR because such values are of higher precision owing to longer counting times. Values are decay-corrected for midpoint of sampling period. Total errors are estimated from counting errors associated with samples, standards, and background as well as from weighing errors. Values preceded by < indicate limits of detection. The identity of the short-lived fission product ¹⁴¹Ce was confirmed by periodic counting during several weeks.

for ⁹⁵Zr, ⁹⁵Nb, ¹⁴¹Ce and ¹⁴⁴Ce which were either not detectable or present in very low amounts in the last particulate sample collected after 15 May. This is not particularly surprising because these radionuclides are chemically reactive and rapidly become associated with particulate matter in sea water^{15,16}. Studies of fallout⁸ in the early 1960s identified the presence of the same fission products ⁹⁵Zr, ⁹⁵Nb, ¹⁴¹Ce and ¹⁴⁴Ce in sea cucumbers at 2,800 m in the northeastern Pacific. Because of the relatively short physical half-lives of these radionuclides, their rapid transit from surface to abyssal depths could not be explained by Stokesian settling models for fallout particles. Thus, it was hypothesized⁸ that zooplankton ingesting contaminated fallout particles and phytoplankton in the surface layers produced faecal pellets which, because of their greater size and density, could transport the radionuclides to depth on time scales of several days instead of the years suggested by Stokes' law. Recent findings of artificial radionuclides in particulate material from deep sediment traps⁵⁻⁷ and in natural zooplankton faecal pellets^{7,17} themselves have lent support to this hypothesis; however, a definite cause and effect relationship has yet to be shown.

Table 2 Chernobyl fallout radionuclides in zooplankton and their faecal pellets

Radionuclide	Concentration (Bq per g dry weight)	
	Zooplankton	Faecal pellets
⁹⁵ Zr	ND	1.4 ± 0.8
⁹⁵ Nb	0.012 ± 0.003	ND
¹⁰³ Ru	0.28 ± 0.06	16.0 ± 1.9
¹⁰⁶ Ru	0.07 ± 0.04	5.8 ± 2.9
¹³⁴ Cs	0.022 ± 0.006	3.4 ± 0.6
¹³⁷ Cs	0.034 ± 0.007	6.3 ± 1.0
¹⁴¹ Ce	0.02 ± 0.01	0.9 ± 0.4
¹⁴⁴ Ce	0.10 ± 0.05	2.5 ± 1.3

Zooplankton was collected on 6 May 1986 at the trap site off the northwestern coast of Corsica by making 15-min oblique tows with a plankton net (333- μ m mesh aperture) between 20 m and the surface. Collections comprised almost exclusively adult copepods of the species *Centropages typicus* (97%), *Acartia clausii* (1%) and *Clausocalanus* sp. (1%). Copepods were immediately transferred to specially designed containers¹⁸ filled with filtered sea water and left to defecate for ~3 h, after which time the faecal pellets were gathered from a small trap at the bottom of each container. All samples were counted at ILMR. Total errors are estimates based on counting statistics of samples and standards and weighing errors. The concentrations of some radionuclides were either below detection limits (ND) or had large errors associated with them because of the small sample size (42.5 mg of faecal pellets) or insufficient counting times or both.

On 6 May zooplankton was sampled from the surface layers at the trap site and live individuals were immediately placed in special devices¹⁸ to collect faecal pellets. Radionuclides in the zooplankton and their faecal pellets (Table 2) closely reflected the distributions seen in the particulate samples (Table 1). Most noteworthy is the close similarity between the concentrations of radionuclides in faecal pellets and those in the particles carrying the radioactivity pulse to 200 m (sample 5 in Table 1). Microscopic examination of the trap samples confirmed that they were rich in faecal material typical of that from copepods. Numbers of intact faecal pellets were in the range 6,000–30,000 per sample, with pellet volumes averaging between 0.4 and 2.4 × 10⁷ μ m³, depending upon the shape of the pellet. Considering volume measurements of the different shape classes of faecal pellets¹⁹ in the material with known pellet densities²⁰, we calculated that zooplankton faecal pellets made up ~70% of the total dry weight of trap sample 5. Moreover our computed best estimates of the particles' sinking speeds (~19–57 m d⁻¹), based on arrival times of peak fallout at the sea surface and at 200 m, are in the range determined for zooplankton faecal pellets^{1,2,18,20}.

The data in Table 2 also indicate that these faecal pellets are substantially enriched in radionuclides compared to the zooplankton producing them. Laboratory studies^{21,22} have shown that Ru and Ce ingested with food are little assimilated by crustacean zooplankton and are primarily excreted with the faeces. In the case of Ce nearly all the radionuclide ingested by euphausiids is defecated within a few hours²².

Although faecal pellets were a large fraction of the sinking particulate matter at our site, this is not always the case. Some sediment trap experiments^{19,23,24} have shown that faecal pellets, particularly in the deep sea, account for <10% of the mass flux. In these cases, most of the flux arrives as large, amorphous aggregates (marine snow) which consist of inorganic and organic detrital particles including faecal matter. In our trap sample 5 the residual material (~30% of total) was an uncharacterized, mucoid floc which closely resembled marine snow. This floc presumably participated in the removal of radioactivity from the overlying waters; however, its relative contribution to radionuclide flux cannot be precisely determined from our radioanalysis of total particles. Nevertheless, marine snow aggregates generally contain lower concentrations of trace elements^{2,25} and fallout radionuclides (S.W.F., S.B. & L.F. Small, in preparation) than faecal pellets and they sink more slowly^{1,2,24}, suggesting that such aggregates are of minor importance in the radionuclide flux we measured.

Radionuclide activity ratios for the samples, including filtered air and unfiltered surface sea water from the same region, are

Table 3 Selected activity ratios of Chernobyl-derived radionuclides in samples from the northwestern Mediterranean

Sample	Collection date	Radionuclide activity ratio		
		$^{137}\text{Cs}/^{141}\text{Ce}$	$^{103}\text{Ru}/^{141}\text{Ce}$	$^{103}\text{Ru}/^{137}\text{Cs}$
Air	3 May	58	110	1.9
Unfiltered sea water	7 May	11	8.0	0.73
Zooplankton	6 May	1.7	14	8.2
Fresh faecal pellets	6 May	7.0	18	2.5
Trapped particles at 200 m	11 May	0.27	1.8	6.8

Air-filter¹⁴ (N. E. Whitehead, S.B. & E.H., in preparation) and sea-water samples (S.B. *et al.*, in preparation) are taken from a coastal station near Monaco ~77 nautical miles northwest of trap site. Radionuclides ^{103}Ru and ^{141}Ce which have short half-lives, have been corrected to 6 May.

Table 4 Vertical fluxes of Chernobyl-derived radionuclides in the northwestern Mediterranean in 1986

Sample	Period	Mass flux (mg m ⁻² d ⁻¹)	⁹⁵ Zr	⁹⁵ Nb	¹⁰³ Ru	Radionuclide		¹³⁷ Cs	¹⁴¹ Ce	¹⁴⁴ Ce
						¹⁰⁶ Ru	¹³⁴ Cs			
						(Bq m ⁻² d ⁻¹)				
1	13–20 April	213.7	}	Before Chernobyl accident						
2	20–26 April	111.5								
3	26 April–2 May	63.9								
4	2–8 May	65.5	—	—	—	—	—	0.0096	—	—
5	8–15 May	53.6	1.31	1.70	1.26	0.29	0.11	0.2035	0.68	0.73
6	15–21 May	57.6	—	—	0.81	0.20	0.11	0.2305	0.063	—

Fluxes through 200 m are shown, as determined by particle interceptor trap measurements. Radionuclide concentrations in the particles are given in Table 1. —, Not detectable.

given in Table 3. It is apparent that in some cases Ce, Ru and Cs are strongly fractionated after entry into the sea and association with particulate matter is most probably affected by their different chemical properties. This is most evident for the lanthanide Ce whose concentration relative to Cs changes dramatically in passing from airborne fallout to particulate matter at 200 m. The decrease in the $^{137}\text{Cs}/^{141}\text{Ce}$ ratio from 7.0 in freshly excreted faecal pellets to 0.27 in the trapped particulates suggests that preferential scavenging of ^{141}Ce occurred as the pellets sank through the upper 200 m. The $^{103}\text{Ru}/^{141}\text{Ce}$ ratios reveal a similar but less marked pattern of fractionation. Such strong scavenging of the Ce radionuclides from the surface layers by biogenic debris is consistent with the reported geochemical behaviour of cerium in the western Mediterranean²⁶ and elsewhere²⁷. Nuclides ^{95}Zr and ^{95}Nb , which were at low concentrations or were undetectable in faecal pellets (Table 2), also seemed to be taken up by the particles during their descent to 200 m (Table 1).

The radionuclide fluxes through 200 m are given in Table 4. For Ce, Zr and Nb most of the flux occurred during the period 8–15 May; on the other hand, the vertical flux of Cs and Ru probably continued for some time after 21 May. For example, during the period 26 April–21 May an inventory of 3.1 Bq m^{-2} can be computed for the long-lived radionuclide ^{137}Cs passing 200 m. This is only a small fraction (~0.2%) of the $1,383 \text{ Bq m}^{-2}$ Chernobyl-derived ^{137}Cs reported to have been deposited as wet and dry fallout in this region between 29 April and 31 May¹⁴. It therefore seems that most of the ^{137}Cs inventory remained in the upper water column after our study. Comparable unpublished total deposition measurements for ^{141}Ce and ^{144}Ce from the Monaco station show inventories of 53 and 41 Bq m^{-2} , respectively, for the same period (S.B., E.H. and N. E. Whitehead, in preparation). Comparing these values with the data in Table 4 suggests that, in sharp contrast to Cs, ~10% of the total Ce deposited in this region of the Mediterranean had fluxed through 200 m as large particles by 21 May. However, given the likelihood of some intraregional variation in fallout deposition, these percentages must be treated as first order estimates.

Radioactive fallout from the Chernobyl accident offered an opportunity to study the processes controlling the vertical transport of surface-introduced contaminants in the sea. We have shown that time-series sediment traps set for relatively high resolution sampling are useful for determining the average sinking speed of contaminated particles. Furthermore, this seren-

dipitous experiment has confirmed the importance of zooplankton faecal pellets in removing artificial radionuclides from surface layers and transporting them rapidly to depth.

The International Laboratory of Marine Radioactivity (ILMR) operates under an agreement between the International Atomic Energy Agency and the Government of the Principality of Monaco. We also thank the Musée Océanographique at Monaco for support and J. La Rosa, S. Heussner and J. Farber-Lorda for technical assistance. This research was supported in part by the Centre National de la Recherche Scientifique, the Commissariat à l'Energie Atomique and grants from the French Flux de Matière dans l'Océan (DYFAMED) and US National Science Foundation VERTEX programmes.

Received 21 April; accepted 8 July 1987.

- Angel, M. V. in *Flows of Materials and Energy in Marine Ecosystems* (ed. Fasham, M. J. R.) 475–516 (Plenum, New York, 1984).
- Fowler, S. W. & Knauer, G. A. *Prog. Oceanogr.* **16**, 147–194 (1986).
- Knapp, A. H., Binkley, K. S. & Deuser, W. G. *Nature* **319**, 572–574 (1986).
- Coale, K. H. & Bruland, K. W. *Limnol. Oceanogr.* **30**, 22–33 (1985).
- Bacon, M. P., Huh, C.-A., Fleer, A. P. & Deuser, W. G. *Deep Sea Res.* **32**, 273–286 (1985).
- Livingston, H. D. & Anderson, R. F. *Nature* **303**, 228–231 (1983).
- Fowler, S. W., Ballestra, S., La Rosa, J. & Fukai, R. *Deep Sea Res.* **30**, 1221–1233 (1983).
- Osterberg, C., Carey, A. C. Jr & Curl, H. Jr *Nature* **200**, 1276–1277 (1963).
- Monaco, A. *et al. Mitt. geol.-paläont. Inst. Univ. Hamburg* (in the press).
- Knauer, G. A., Karl, D. M., Martin, J. H. & Hunter, C. N. *J. mar. Res.* **42**, 445–462 (1984).
- Webb, G. A. M., Simmonds, J. R. & Wilkins, B. T. *Nature* **321**, 821–822 (1986).
- Thomas, A. J. & Martin, J. M. *Nature* **321**, 817–819 (1986).
- Summary Report on the Post-Accident Review Meeting on the Chernobyl Accident. Safety Series No. 75-INSAG-1* (IAEA, Vienna, 1986).
- Ballestra, S., Holm, E., Walton, A. & Whitehead, N. E. *J. env. Radioactivity* **5** (in the press).
- Burton, J. D. in *Chemical Oceanography* 2nd edn **3**, 91–191 (eds Riley, J. P. & Skirrow, G.) (Academic, London, 1975).
- Coughtrey, P. J. & Thorne, M. C. *Radionuclide Distribution and Transport in Terrestrial and Aquatic Ecosystems. A Critical Review of Data Vol. 1* (Balkema, Rotterdam, 1983).
- Higgo, J. J. W., Cherry, R. D., Heyraud, M. & Fowler, S. W. *Nature* **266**, 623–624 (1977).
- Small, L. F., Fowler, S. W. & Unlu, M. Y. *Mar. Biol.* **51**, 233–241 (1979).
- Heussner, S., Monaco, A. & Fowler, S. W. *Mitt. geol.-paläont. Inst. Univ. Hamburg* (in the press).
- Komar, P. D., Morse, A. P., Small, L. F. & Fowler, S. W. *Limnol. Oceanogr.* **26**, 172–180 (1981).
- Keckes, S., Fowler, S. W. & Small, L. F. *Mar. Biol.* **13**, 94–99 (1972).
- Fowler, S. W., Heyraud, M., Small, L. F. & Benayoun, G. *Mar. Biol.* **21**, 317–325 (1973).
- Pilskaln, C. H. & Honjo, S. *Global biogeochem. Cycles* **1**, 31–48 (1987).
- Asper, V. L. *Deep Sea Res.* **34**, 1–17 (1987).
- Hebel, D., Knauer, G. A. & Martin, J. H. *J. Plankton Res.* **8**, 819–824 (1986).
- Risler, N., Lambert, C. E., Buat-Menard, P. & Chesselet, R. in *VI^e Journées d'Etudes sur les Pollutions Marines en Méditerranée*, 119–128 (Commission Internationale pour l'Exploration de la Mer Méditerranée, Monaco, 1985).
- Elderfield, H. & Greaves, M. J. *Nature* **296**, 214–219 (1982).
- Brouard, J. *Mém. Inst. océanogr. Monaco* **3**, 1–20 (1971).
- Yokoyama, Y. & Nguyen, H. V. in *Isotope Marine Chemistry* (eds Goldberg, E. D., Horibe, Y. & Saruhashi, K.) 259–289 (Rokkakuhō, Tokyo, 1980).

Gravitational haemodynamics and oedema prevention in the giraffe

Alan R. Hargens*¶, Ronald W. Millard†, Knut Pettersson‡ & Kjell Johansen§||

* Division of Orthopaedics and Rehabilitation (V-151), University of California, San Diego and Veterans Administration Medical Center, San Diego, California 92161, USA

† Department of Pharmacology and Cell Biophysics, University of Cincinnati, Ohio 45267, USA

‡ Department of Cardiovascular Pharmacology, AB Hässle, S-431 83 Mölndal, Sweden

§ Department of Zoophysiology, University of Aarhus, DK-8000 Aarhus C, Denmark

Because it is so tall, the giraffe, *Giraffa camelopardalis*, provides an important animal model for investigating adaptive mechanisms to orthostatic (gravitational) pressure changes. Previous physiological studies of the giraffe have concentrated on arterial blood pressures in the heart and neck¹⁻³. Briefly, these investigations revealed that arterial pressure near the giraffe heart is about twice that in humans, to provide more normal blood pressure and perfusion to the brain. Another important question is that of how giraffes avoid pooling of blood and tissue fluid (oedema) in dependent tissues of their extremities. As monitored by radiotelemetry, the blood and tissue fluid pressures that govern transcapillary exchange vary greatly with exercise. These pressures, combined with a tight skin layer, move fluid upward against gravity. Other mechanisms that prevent oedema include precapillary vasoconstriction and low permeability of capillaries to plasma proteins.

The forces that regulate transcapillary fluid balance were first identified by the pioneering British physiologist Ernest Starling⁴. These fluid pressures are commonly called 'Starling pressures' and include capillary blood pressure P_c , interstitial fluid pressure P_i , blood colloid osmotic pressure π_c and interstitial fluid colloid osmotic pressure π_i . The accepted formula for transcapillary fluid flow J_c is:

$$J_c = L_p A [(P_c - P_i) - \sigma_p (\pi_c - \pi_i)] \quad (1)$$

where L_p is fluid conductivity, A is capillary membrane surface area, and σ_p is the protein reflection coefficient for the capillary membrane⁵.

During the 1985 Giraffe Physiology Expedition to Africa, eight male and female giraffes 3-4 m tall (518 ± 131 kg body weight) were studied in terms of their Starling pressures, regional blood pressures and flows, and histomorphology. These initial studies of the giraffes in a stationary upright posture were possible using a mixture of detomidine ($23 \mu\text{g kg}^{-1}$) and azaparone ($350 \mu\text{g kg}^{-1}$) to sedate the giraffes during all catheterizations and blood flow measurements⁶. Arterial and venous blood pressures in the neck and foot were determined by saline-filled polyethylene tubing connected to pressure transducers kept level with each catheter tip. This also allowed periodic sampling of arterial and venous blood for estimating colloid osmotic pressure of capillary blood π_c (ref. 7). Concurrently, interstitial fluid pressure P_i was measured by the wick catheter technique⁸, maintaining the pressure transducer level with the catheter tip. Empty wick catheters were employed to collect 5- μl samples of interstitial fluid for determination of colloid osmotic pressure⁸.

Jugular vein pressures were measured as a function of hydrostatic height in three giraffes using the Millar Mikrotip transducer, in which the pressure-sensing surface was at the tip of

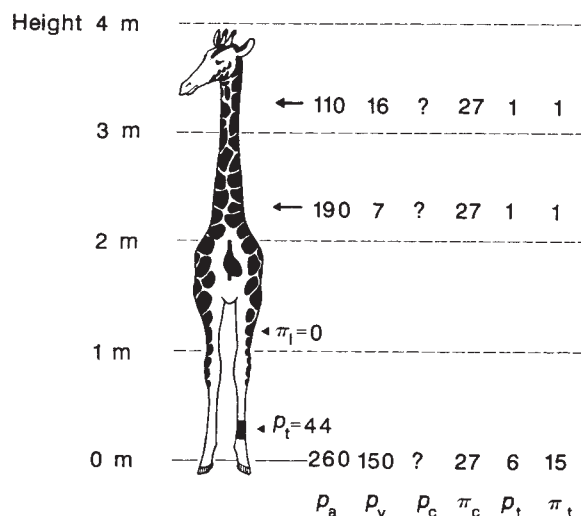


Fig. 1 Mean arterial P_a , venous P_v and transcapillary fluid pressures (P_c , π_c , P_t , and π_t) in the giraffe at hydrostatic levels between the head and feet during upright, standing posture. Symbols as in equation (1). Colloid osmotic pressure of capillary blood π_c was estimated by averaging arterial and venous samples. All pressures are in mm Hg. Lymph samples obtained from the leg had only trace amounts of protein ($\pi_t = 0$, $N = 2$) and foot samples for π_t were often contaminated by blood and therefore were less reliable than those at neck levels. Note that P_t beneath the tight skin and fascia of the legs ranged between 40 and 50 mm Hg ($N = 6$), indicating the presence of a natural antigravity suit in the giraffe.

the catheter. These measurements were correlated with studies of venous valve spacing in dissected veins. Subsequently, a Konigsberg radiotelemetry system was mounted in a backpack at the base of the neck of four giraffes. This allowed continuous monitoring of blood and interstitial fluid pressures for one day and night during which the fully alert giraffe was free to move.

Local blood flows in neck and leg tissues were measured by the ¹³³Xe washout procedure⁹. Because of local culling programmes, two giraffes were killed, allowing collection and histomorphometry of multiple tissue samples from various regions of their bodies.

In the upright standing posture, mean values for arterial pressures qualitatively matched the expected gravitation pressure gradient between the head and foot (Fig. 1). A gravitational pressure gradient starting at heart level also accounted for pressure in the digital vein of the forelimb. Although P_c was not measured directly in the giraffe, it is probably near P_v in the feet (150 mm Hg) and near P_v at the top of the neck (10-20 mm Hg). Surprisingly π_c was identical to that in humans⁸ and therefore, π_c in the giraffe foot ($N = 9$ samples in different feet of 7 giraffes) offers no unusual resorptive pressure for preventing dependent oedema. Although some P_t values in the neck were negative, average P_t ranged between 1 and 6 mm Hg ($N = 11$), except under the tight skin and fascia of the extremities where mean P_t was 44 mm Hg ($N = 6$). Interestingly, π_t was very low (1 mm Hg, $N = 7$ for upper neck and $N = 7$ for lower neck), except in foot samples that were often contaminated by blood. This finding provided evidence that giraffe capillaries are highly impermeable to plasma proteins and that σ_p approximated unity. This conclusion was supported by studies of peripheral lymph that indicated only trace amounts of protein were present and lymph colloid osmotic pressure π_t equalled zero ($N = 2$).

Inserting our values for the Starling pressures into equation (1) yields a net resorptive pressure of -7 mm Hg in the giraffe neck and net filtration pressures of +88 to +152 mm Hg for

|| Deceased 5 March 1987.

¶ Present address: Space Physiology Branch, NASA-Ames Research Center, Moffett Field, California 94035, USA.

tissues of the leg. These calculations suggest that giraffes are susceptible to dependent oedema in an upright, stationary posture. However, our radiotelemetry data ($N = 8$ studies in four giraffes) for the giraffe foot indicate that P_a (ranging between 70 to 380 mm Hg), P_v (-250 to +240 mm Hg) and P_t (-120 to +80 mm Hg) were highly variable during normal walking. Consequently, it appears there is an effective pumping mechanism in the vascular and interstitial spaces for moving blood and interstitial fluid against gravity. The tight skin and fascial layers of the giraffe leg provide a functional 'antigravity suit' to prevent pooling of blood and interstitial fluid in dependent tissues.

The pressure gradient down the jugular vein was about one-tenth of, and in the opposite direction to, that expected for a standing column of blood (Fig. 2). These venous pressures are interesting and deserve further investigation. A siphon-like model for circulations above heart level has been suggested^{10,11} but our arterial and venous pressure results in the giraffe do not support such a model.

Blood flows in skeletal muscle of the neck and the leg both averaged 4 ml min⁻¹ per 100 g ($N = 3$ studies in each region). Therefore, it was apparent that arteriolar smooth muscle constriction (tonus) was effective in normalizing blood flow in the leg despite significantly higher perfusion pressure. Vascular resistance in leg skeletal muscle was calculated to be twice that in neck skeletal muscle. Pronounced arterial and arteriolar wall hypertrophy in dependent tissues was a prominent feature of our giraffe histomorphometric studies¹².

The giraffe is certainly unique in its adaptation to high gravitational pressures in its cardiovascular system. The oedema-preventing mechanisms that were detected in giraffe legs included variable and sometimes negative P_v and P_t ; low permeability capillary membranes to retain intravascular proteins; arterial wall hypertrophy and precapillary vasoconstriction to reduce dependent blood flow; a prominent lymphatic system; and finally a skin and fascial antigravity suit combined with one-way valves in the veins and lymphatics to reduce venous capacitance and to propel blood and peripheral lymph upwards against a gravitational pressure gradient. This study demonstrates that intravascular and interstitial fluid pressures are highly variable during normal exercise and that studies of recumbent and upright and stationary animals may give misleading information about transcapillary fluid balance. Therefore, normal exercise activities with concomitant pumping of veins, interstitial fluids and peripheral lymph are important for oedema prevention in dependent tissues. That P_a and P_t were so much less than zero during exercise surprised us and should be investigated in future studies. Note that taller mammals (such as giraffes and humans) have relatively thick fascial layers around their lower extremities whereas shorter mammals (such as rats and rabbits) do not. Apparently the cyclical compression and decompression associated with muscular activity and tight fascial enclosures¹³ explain the highly variable P_a , P_v , and P_t that were observed.

From our ¹³³Xe washout results, it is apparent that reflex precapillary vasoconstriction normalizes blood flow in leg muscle (compared to neck muscle) and may reduce capillary filtration area A in dependent tissues, as is observed in humans during quiet standing^{9,14}. This oedema-preventing mechanism may have been underestimated in our studies because all giraffes were sedated during measurements of blood flow. Also during upright posture, colloid osmotic pressure increases rapidly in foot veins of humans¹⁵, another mechanism for preventing oedema. Because we observed no elevation of π_c in dependent vessels, either this mechanism was not detected during giraffe sedation or it was not present.

Some anatomical adaptations of the giraffe and human obviously represent developmental adjustments to high and variable gravitational pressures. For example, the relatively impermeable capillaries in the giraffe ($\sigma_p \approx 1$) tends to raise π_c and lower π_t , π_i and P_t . Capillary basement membrane thickness

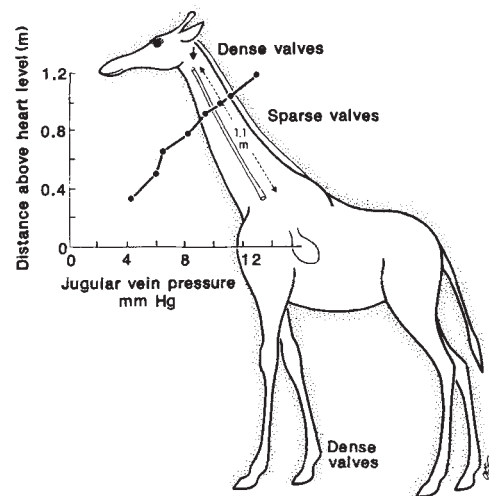


Fig. 2 Pressure as a function of hydrostatic height in the jugular vein and intervalve spacing in the giraffe neck and leg. Solid arrow, catheter insertion point; dotted arrows, maximum extension of Millar tip. In one giraffe the jugular vein pressure gradient was ~9 mm Hg per 90 cm of vertical distance or 0.14 cm H₂O per cm of vertical distance. Intervalve distances were short in the head, distal neck and distal leg but long in the proximal neck. The existence of closely spaced valves in the head and distal neck as compared to sparse valves in the proximal neck indicates their importance for preventing retrograde venous flow, for example, during short periods when the giraffe's head is lowered below heart level during drinking. Unfortunately, the Millar transducer was insufficiently long to measure intrathoracic venous pressure. These data indicate that P_v in the jugular vein is opposite to that predicted by gravity and may suggest compartmentalization of the blood column.

increases twofold between neck muscle and leg muscle of adult giraffes and humans¹⁶. On the other hand, such membranes in the human fetus are uniform and considerably thinner than those in children and adult humans.

This research was supported by grants from NIH, NSF, and the National Geographic Society. We thank our hosts Professor Wouter van Hoven and the Department of Zoology, University of Pretoria for providing facilities. We also thank the following for assistance: Dr Richard Burroughs and the National Zoological Gardens in Pretoria, Dr D. G. A. Meltzer and Roodeplaat Research Labs, Dr Frik J. Stegmann and University of Pretoria Veterinary School at Onderstepoort, Dr David H. Gershuni, Orthopaedic Division, Veterans' Administration Medical Center and University of California at San Diego, Dr Paul Calle and Cheyenne Mountain Zoological Park at Colorado Springs, and the Department of Physiology, University of Pretoria.

Received 27 May; accepted 2 July 1987.

- Goetz, R. H. *et al. Circulation Res.* **8**, 1049-1058 (1960).
- Van Citters, R. L., Kemper, S. & Franklin, D. L. *Comp. Biochem. Physiol.* **24**, 1035-1042 (1968).
- Patterson, J. L. Jr *et al. Ann. N.Y. Acad. Sci.* **127**, 393-413 (1975).
- Starling, E. H. *J. Physiol., Lond.* **19**, 312-326 (1896).
- Hargens, A. R. in *Handbook of Bioengineering* (eds Skalak, R. & Chien, S.) Ch. 19, 1-25 (McGraw-Hill, New York, 1987).
- Meltzer, D. G. A. *et al. in Exotic Animals in Research* (ed. Dawson, P.) 42-46 (S. Afr. Assoc. Lab. Animal Sci., Pretoria, 1985).
- Aukland, K. & Johnson, H. M. *Acta physiol. scand.* **90**, 485-490 (1974).
- Hargens, A. R. *et al. Microvasc. Res.* **22**, 177-189 (1981).
- Henriksen, O. *Acta physiol. scand. Suppl.* **450**, 1-48 (1977).
- Burton, A. C. *Physiology and Biophysics of the Circulation* 2nd edn, 97-103 (Year Book, Chicago, 1972).
- Badeer, H. S. *Comp. Biochem. Physiol.* **83A**, 207-211 (1986).
- Petersson, K. *et al. Proc. XXX Congress, IUPS XVI*, P414.16 (1986).
- Sejersted, O. M. *et al. J. appl. Physiol.* **56**, 287-295 (1984).
- Mellander, S., Öberg, B. & Odelram, H. *Acta physiol. scand.* **61**, 34-48 (1964).
- Noddeland, H., Aukland, K. & Nicolaysen, G. *Acta physiol. scand.* **113**, 447-454 (1981).
- Williamson, J. R., Vogler, N. J. & Kilo, C. *Am. J. Pathol.* **63**, 359-370 (1971).

Treponemal infection in a Pleistocene bear

Bruce M. Rothschild & William Turnbull*

Department of Rheumatology, St Elizabeth Hospital Medical Center, Northeastern Ohio Universities College of Medicine, Youngstown, Ohio 44 501-1790, Kansas University Museum of Natural History, Lawrence, Kansas 66044 and Carnegie Institute, Pittsburgh, Pennsylvania 15213, USA

*Department of Vertebrate Paleontology/Geology Field Museum of Natural History Chicago, Illinois 60615, USA

The age and origins of the organisms that cause syphilis (treponemes) have long been matters for controversy¹⁻³. The widely-held belief that Columbus's ship brought the disease from the New World to Europe rests on identification of the classic lesions in Inca, Aztec and Mississippian bones that date from 1,000 to 3,000 years before present³. But these were not confirmed by immunological techniques. We have observed lesions characteristic of treponemal infection⁴ in a Pleistocene bear from Indiana (dated 11,500 years BP) which give a positive result when tested with the antisera used by the US Center for Disease Control for verification of syphilis infection. This is the earliest detection of treponemal disease using contemporary techniques.

Gummatous lesions, associated with spiculated periosteal reaction, were noted in the mandible, humeri, radii, ulna and thoracic vertebrae (T3-T5) (Figs 1 and 2) of an Indiana Pleistocene bear, *Arctodus simus* (Chicago Field Museum of Natural History, PM24880), carbon-13 dated at 11,500 ± 520 years. The observed pathology (Figs 1 and 3) was typical and highly characteristic of that reported in venereal syphilis⁴⁻⁶. Absence of sequestra and presence of destructive areas, surrounded by reactive sclerosis, argue against a diagnosis of pyogenic or tubercular infection^{2,6}. Radiological examination confirmed the spiculated nature of the periosteal reaction and absence of sequestra (Fig. 2). Subjecting sections of involved thoracic vertebra to direct immunofluorescent analysis, using anti-syphilis, *Treponema phagedenis*-absorbed conjugate number 0099 (Center for Disease Control), revealed focal clumps of treponemal antigen in the margins of the vertebral erosion, but nowhere else in the vertebral matrix. Use of absorbed conjugate and conjugates to *Neisseria gonorrhea*, *Streptococcus*, and to *Legionella pneumophila* revealed no such fluorescence.

The polyclonal antiserum was chosen to test the treponemal hypothesis as it is extremely well characterized and, being polyclonal, will bind to the full spectrum of treponemal antigens. It is impossible to predict which treponemal antigen(s) might survive the millennia sufficiently intact to be detected immunologically, and this antiserum is sufficiently specific that using a monoclonal antibody is unlikely to decrease the probability of cross reaction.

Immunofluorescent histological identification of treponemal antigen in contemporary histological sections occasionally shows intact organisms, but more typically large clumps of fragmented organisms or free treponemal antigen^{7,8}, identical in appearance to those noted in our specimen of *Arctodus simus*, are seen.

The only known treponemal bone diseases are venereal syphilis, endemic syphilis and yaws. Although 65% of syphilitic individuals do not show anatomical evidence of syphilis at autopsy², bone lesions are reported in 9.5-23.6% of patients with acquired syphilis^{2,9}, 3-5% with endemic syphilis^{2,10} (Bejel) and 10-15% of yaws^{2,11}. Gummatous osteitis, periostitis, and draining sinuses are the most common lesions in both syphilis and yaws^{2,12-14}. Distribution is predominantly to weight-bearing joints in syphilis⁹, and to the forelimbs in yaws^{2,11}. Joints specifically susceptible to trauma are more often affected in both^{2,9,11}.

Fig. 1 The fossil humerus from *Arctodus simus* revealing periosteal reaction. Magnification $\times 0.45$. Scale bar, 5 cm.



Fig. 2 Thoracic vertebrae from *Arctodus simus*. Erosive spondylitis with exuberant new bone formation. Magnification $\times 0.7$

Gross examination of affected bones reveals spiculation and avulsion-type diaphyseal exostoses^{2,10,11,15} of intact periosteum, also seen in the *Arctodus* specimen; the subperiosteal callous formation seen is also characteristic of human syphilis. Freedman and Menshan indicated that the spine was often the site of invasion of periosteum and bone biopsy productive for identification of *T. pallidum*¹⁶.

The spinal lesions in syphilis and yaws are indistinguishable, characteristically lytic anteriorly with hyperostoses. Typically two to three adjacent vertebrae are involved and the anterior and lateral longitudinal ligaments are often calcified^{11,16,17}. The

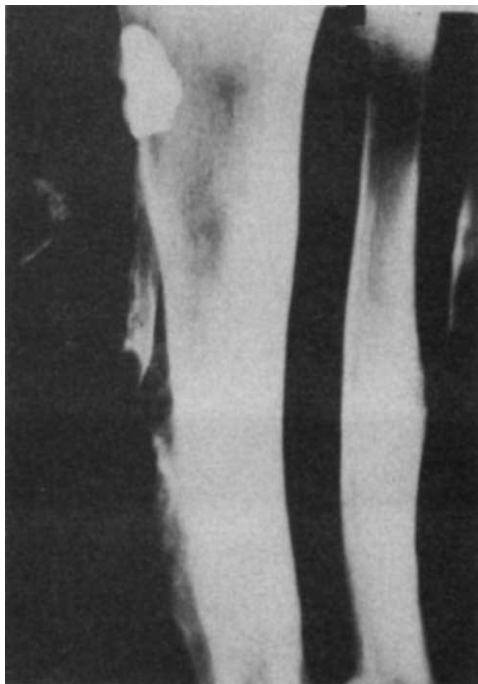


Fig. 3 Radiograph of the *Arctodus simus* humerus shown in Fig. 1. Magnification $\times 0.4$.

disk spaces may or may not be retained and the vertebral endplates take on a moth-eaten appearance¹⁷ (Fig. 2).

As yaws is the only treponeme known to have an animal reservoir (cynocephal African monkeys¹⁸, and as only pinta and yaws have known insect vectors (*Simulium* and *Hippelates pallipes* (a fly) respectively^{19,20}), yaws is the most probable cause of this disease. Early accounts suggesting that yaws was not present in the New World may have confused the skeletal alterations with those characteristic of syphilis. Observations of treponemal disease in pre-Columbian Indians are equally difficult to assess. Because lesions of congenital syphilis are notable by their absence in the pre-Columbian New World, yaws or endemic syphilis are more probable as causes than venereal syphilis². One could also speculate that an extinct treponeme¹⁸ is responsible.

Presence of a treponemal infection in a Pleistocene bear indicates a long residence time for this infection in the western hemisphere and suggests that either yaws or syphilis or both could reasonably be expected in pre-Columbian New World Indians, as has previously been suggested. It is interesting that good pre-Columbian osteological evidence tends to be lacking outside the New World, although some question arises with respect to the diagnosis of leprosy in the Middle Ages. It will be of interest to pursue similar immunological studies of other pre-Columbian anthropological specimens.

We thank Drs Larry Martin and Gordon Bell, respectively for manuscript review and critical commentary; Dr Jerry Volenac for logistical support; and Dr Elizabeth Hunter for providing conjugate and technical consultation.

Received 21 April; accepted 16 July 1987.

1. Henshen, F. in *The History and Geography of Diseases* 123–124, 131, 133 (Delacorte, New York, 1962).
2. Steinbock, R. T. in *Paleopathologic Diagnosis and Interpretation*, 89–169 (Thomas, Springfield, 1976).
3. Moller-Christensen V. *Abbotempo* 1, 20–25 (1969).
4. Ortner, D. J. & Putschar, W. G. in *Identification of Pathologic Conditions in Human Skeletal Remains*, 180–218 (Smithsonian Institution, Washington DC, 1981).
5. Kampmeier, R. H. in *Essentials of Syphilology* (Lippincott, Philadelphia, 1943).
6. Resnick, D. & Niwayama, G. in *Diagnosis of Bone and Joint Disorders*, 2154–2237 (Saunders, Philadelphia, 1981).
7. Yobs, A. R., Brown, L. & Hunter, E. F. *Arch. path. Lab. Med.* 77, 220–225 (1964).
8. Handsfield, H. H., Lukehart, S. A., Sell, S., Norris, S. J. & Holmes, K. K. *Arch. Derm.* 119, 677–680 (1983).

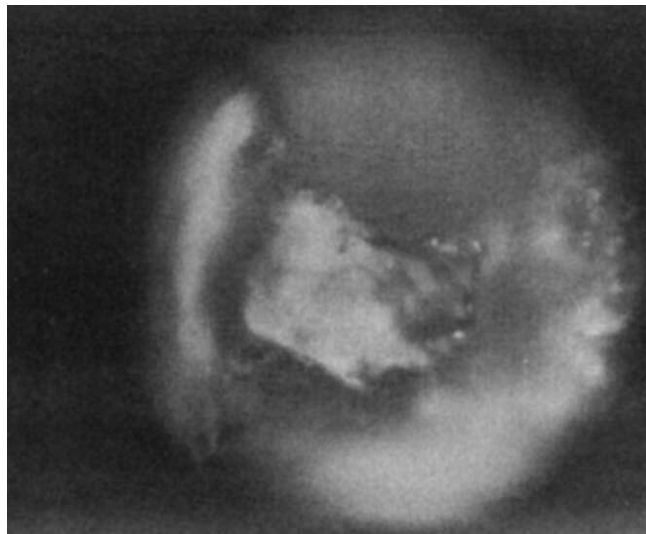


Fig. 4 Cross-section of gumma-bone interphase in *Arctodus simus*. Focal green immunofluorescence with *T. pallidum*-specific *T. phagedenis*-absorbed conjugate. Magnification $\times 16$.

Methods. The fourth thoracic vertebrae from *Arctodus simus* and representative fossils from *Bison bison*, *Ovis* and *Teleocerus* (with various forms of infectious lesions) were sectioned using a diamond saw. Sections (~ 3 cm square) were subjected to epi-immunofluorescent microscopy using 1:10 and 1:20 dilutions of *T. phagedenis*-absorbed or high titre *T. pallidum*-absorbed antisyphilis conjugate number 0099 and conjugate to *Neisseria gonorrhea*, *Streptococcus* and to *Legionella pneumophila* (all provided by the Center for Disease Control). Conjugate number 0099 was absorbed with the Reiter treponeme *T. phagedenis* at 1:1 dilution of thioglycolate broth (10% rabbit serum) cultured *T. phagedenis*, eliminating cross-reacting non-pathogenic treponemal and *Borrelia* (for example Lyme) antigens.

9. Whitney, J. L. *J. Am. med. Ass.* 65, 1986–1989 (1915).
10. Hudson, E. H. in *Non-venereal Syphilis: a Sociological and Medical Study of Bejel* (Livingston, London, 1958).
11. Moss, W. L., Bigelow, G. H. *Bull. J. Hopkins Hosp.* 33, 43–47 (1922).
12. Helfet, A. J. *J. Bone Joint Surg.* 26B, 672–681 (1944).
13. Goff, C. W. in *Diseases in Antiquity* (eds Brothwell, D. R. & Sandison, A. T.) 170–187 (Thomas, Springfield, 1967).
14. Dismukes, W. E., Delgado, D. G., Mallnee, S. V. & Myers, T. C. *J. Am. med. Ass.* 236, 2646–2648 (1976).
15. Ziesche, H. *Mitt. Grenzgeb. med. Chir.* 22, 357–389 (1910).
16. Freedman, E. & Meschan, I. *Am. J. Roent.* 49, 756–765 (1943).
17. Horwitz, M. A. *A Rheumat. Dis.* 7, 200–208 (1948).
18. Hackett, C. J. *Bull. WHO* 29, 7–14 (1963).
19. Turner, T. B. *Am. J. Hyg.* 25, 477–506 (1937).
20. Sanchez, F. G., Mazzotti, L. & Salenas, E. G. *Salud Publ. Mex.* 3, 183–188 (1961).

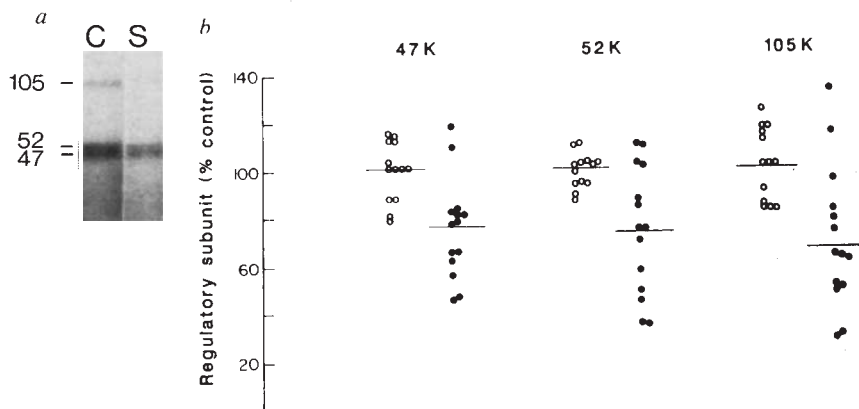
A molecular mechanism for long-term sensitization in *Aplysia*

Steven M. Greenberg, Vincent F. Castellucci,
Hagan Bayley & James H. Schwartz

Howard Hughes Medical Institute, Center for Neurobiology and Behavior, Columbia University, 722 W. 168th Street, New York, New York 10032, USA

Sensitization of the gill- and siphon-withdrawal reflex in *Aplysia* is thought to result from a set of molecular processes with different time courses¹: short-term sensitization is explained by cyclic AMP-dependent modulation of ion-channel function in sensory neurons lasting minutes²; memory that endures for hours or longer, by the expression and distribution within the neurons of new gene products³. Because gene induction and axonal transport are relatively slow, there may also be a need for a distinct form of intermediate

Fig. 1 The effect of long-term sensitization on cAMP-dependent protein kinase R subunits. **a**, Autoradiographs of labelled samples of tissue containing sensory cells with equal amounts of protein, one from a control (C), the other from a sensitized (S) animal. Size markers M_r in thousands. **b**, Summary of results with trained animals. Each point represents a radiolabelled component from a single naive (○) or trained (●) animal. Means are marked by a horizontal line. Values for each species of regulatory subunit in samples from the trained animals were significantly lower than those for the controls (for the M_r 47 K species (left), $P < 0.02$; for the 52K (middle) and 105K (right) species, $P < 0.01$, two-tailed Student's *t*-test). The amounts measured in trained animals varied more than in controls, perhaps reflecting the variation in aptitude regularly observed in behavioural experiments with *Aplysia*.



Methods. *Aplysia californica* (100–200 g, from the Howard Hughes Medical Institute Mariculture Resource Facility at the Woods Hole Oceanographic Institution, Massachusetts) were trained as described⁶; briefly, 13 control and 14 trained animals were studied in a total of three separate experiments. Siphon withdrawal times were significantly greater than their pre-training scores. The middle portion of the left half of the abdominal ganglion, which contains the siphon sensory neurons as well as some interneurons and motor neurons, was dissected out in sea water with elevated Mg^{2+} (220 mM) and reduced Ca^{2+} (1 mM) and transferred to 20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 3 mM EGTA, 0.1% Nonidet P-40, 0.5 mM 2-mercaptoethanol, 5 mM benzamidine, 10 μ g ml⁻¹ aprotinin (230 trypsin inhibitory units mg⁻¹), 20 μ M leupeptin, and 1 mM phenylmethylsulphonylfluoride (PMSF, 50 μ l) (Buffer A), homogenized, and centrifuged for 30 s at 1,000g to remove poorly homogenized material. The supernatant was brought to 1 μ M [³²P]8-N₃cAMP (ICN) and incubated for 90 min at 23 °C, then 10 min at 0 °C followed by exposure for 5 min at 0 °C to ultraviolet light with a wavelength of 302 nm. After photolysis, proteins in the samples were separated by SDS-PAGE and detected by autoradiography⁸. Radiolabelled proteins were quantified by densitometric scanning of the autoradiograms. Values were normalized to the amount of protein applied; protein was measured by staining with Amido black 10B (ref. 25) using bovine serum albumin as standard. Control experiments with added [³²P]8-N₃cAMP showed that cAMP bound to R subunits exchanges completely under the conditions used: subunits were saturated with cAMP, free cAMP was then removed by gel filtration on Sephadex G-50 and the preparation was photolabelled. These samples were labelled to the same extent as untreated aliquots.

memory to bridge the short- and long-term processes⁴. We now report that a protocol producing long-term sensitization results in a decrease in the amount of regulatory subunits of the cAMP-dependent protein kinase in animals 24 h after training, with no effect on the catalytic subunit. The loss appears to be post-translational. Because a decrease in the ratio of regulatory to catalytic subunits would result in elevated kinase activity after cAMP has returned to its unstimulated concentration in sensory cells, it could be the molecular mechanism of intermediate memory.

Activation of cAMP-dependent protein kinase in *Aplysia* sensory neurons produces short-term sensitization, and it has been proposed² that the long-term form of this behavioural modification might be mediated by an altered kinase molecule. The holoenzymes of the kinase, composed of regulatory (R) and catalytic (C) subunits, dissociate in the presence of cAMP, releasing active catalytic monomers. Because the amount of cAMP does not remain elevated in sensory neurons after long-term training⁵, it was suggested that the subunits of this enzyme are modified to be active at pre-stimulation concentrations of cAMP. This change to a more responsive kinase might result either from the induction of new forms of R subunits with higher affinity for cAMP (as originally proposed²) or by decreasing the ratio of R to C subunits (the 'R-to-C ratio').

We measured R and C subunits in tissue samples containing the LE sensory neuron cluster (a group of 24 small identified cells²) dissected from left abdominal hemiganglia of animals 24 h after sensitization. Long-term training consisted of four series of tail shocks delivered during a total period of 90 min (ref. 6). R, the cAMP-binding component of the kinase, was assayed using the cAMP analogue [³²P]8-N₃cAMP. This photoaffinity reagent has previously revealed five distinct subunits in *Aplysia* neurons^{7,8}. The R subunits in *Aplysia* nervous tissue appear as three specifically labelled bands with relative molecular masses (M_r) 47,000 (47K), 52K and 105K on one-dimensional SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 1a). In three separate experiments, training significantly reduced the amounts of R in all the weight classes (Fig. 1b and Table 1), and no new photolabelled components

Table 1 Effects of sensitization training on cAMP-dependent protein kinase and on tubulin

Source of cells	R (Normalized assay values)	C (Normalized assay values)	Tubulin
Control	1 ± 0.03 (19)	1 ± 0.05 (15)	1 ± 0.05 (10)
Sensitized			
Long-term	0.75 ± 0.06 (14)	0.93 ± 0.04 (11)	0.93 ± 0.06 (11)
Short-term	0.97 ± 0.09 (5)	0.99 ± 0.07 (6)	—

R and C subunits of the cAMP-dependent protein kinase and tubulin were assayed in homogenates of cells from control and both long-term and short-term sensitized animals 24 h after they had been trained. Values (mean ± s.e.m.) were normalized by setting control values to 1. **Methods.** Individual homogenates from control and sensitized animals (number shown in parentheses) were photoaffinity labelled with [³²P]8-N₃cAMP and subjected to SDS-PAGE as described in Fig. 1 legend. Relative tubulin content was estimated by scanning the Coomassie-blue-stained gels at 633 nm. The R subunits were estimated by scanning autoradiograms of the same gels. The control value for incorporation of photoaffinity reagent was 5.8 pmol per mg protein. Specific covalent incorporation of [³²P]8-N₃cAMP into extracts of *Aplysia* nervous tissue was half that of noncovalent [³H]cAMP binding¹⁰, the latter measured under conditions in which each mole of R subunit binds 2 moles of cAMP (ref. 20). This result indicates that [³²P]8-N₃cAMP labels R subunits in *Aplysia* (as in other animals^{21–24}) with a stoichiometry of 1:1. An untreated portion of each homogenate was assayed for kinase activity by measuring the rate of phosphorylation of a synthetic peptide substrate (Kemptide, Sigma) in the presence of 10 μ M cAMP, as described previously^{8,10}. The control value for kinase activity was 1.7 nmol ATP per min per mg protein at 23 °C. The values of R from control and long-term-trained animals differ significantly ($P < 0.01$ using Student's *t*-test).

appeared. The loss was restricted to R subunits. The amount of C did not change as a consequence of training (Table 1). We also found no difference in the amounts of tubulin between the samples of neurons from trained and control animals. Further, no differences were apparent in any of the unidentified stained

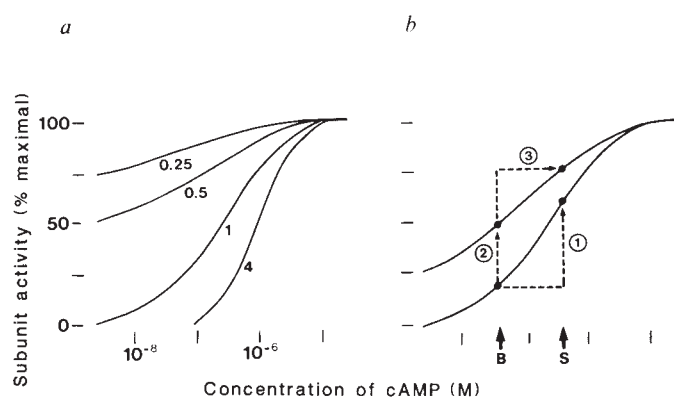


Fig. 2 Model for the effect of short-term and long-term training on phosphorylation by the C subunit. In *a*, each curve is designated according to the stoichiometric ratio of R-to-C subunit that it represents. The curves showing 4:1 and 1:1 ratios are drawn from published results²⁶; the others were calculated assuming that catalytic activity is proportional to the amount of C not associated with R. In *b*, curves for the 1:1 and 0.75:1 ratios are used to illustrate the effect of a 25% loss of R. The actual effects, which would depend on the different affinities of the cAMP-binding sites of the various R subunits¹⁰ as well as the activity of any endogenous kinase inhibitors, have not been determined.

bands. In any particular animal, the decrease in each of the three R-subunit classes was of similar magnitude.

The decrease in R-to-C ratio is specific to long-term sensitization. When animals were stimulated by a brief train of electrical shocks to the tail⁹, a protocol that we confirmed produces only short-term sensitization of the siphon-withdrawal reflex, we found no change in either R or C subunits after 24 h (Table 1). The change in ratio also was not seen in isolated ganglia treated by two kinds of protocols that have been shown to produce the synaptic facilitation underlying short-term sensitization^{2,3}. In the first experiment ($n=5$), nerve stimulation⁹ was used and the R-to-C ratio measured 1 h later in pleural sensory cell clusters: cells from the contralateral ganglia served as controls. In the other experiment, ganglia were exposed to 5-hydroxytryptamine (5HT, serotonin) with no change in the ratio¹⁰.

The effect of decreasing the amount of R relative to C should be to increase kinase activity at any subsaturating concentration of cAMP. The dependence of enzyme activity on cAMP at several different ratios of R to C is shown in Fig. 2*a*: the lower the ratio, the lower the concentration of cAMP needed to activate the kinase. Thus, even the small change in the ratio found after training would be expected to promote phosphorylation of proteins in sensory cells after cAMP returns to the basal concentration. Because increased C-subunit activity facilitates neurotransmitter release from sensory neurons^{11,12}, this might constitute a molecular mechanism for prolonging memory. In sensory neurons of untrained animals, about 20% of the total binding sites on R are occupied by cAMP; short-term training raises this value to 60% (ref. 10) (pathway 1 in Fig. 2*b*). At the lower ratio of R to C found in neurons of long-term trained animals, the concentration of cAMP, which is equivalent to the basal (B) value in the untrained cells⁵, could activate the kinase to an extent similar to that produced by the higher concentration of cAMP (S) formed in response to a short-term stimulus (Fig. 2*b*, pathway 2). Recent evidence indicates that the same K^+ -channel modulated during short-term sensitization by a cAMP-dependent process is also modulated in the long term (ref. 13 and N. Dale, E. R. Kandel and S. Schacher, personal communication). This provides support for the idea that long-term sensitization makes use of the same phosphorylation reaction as the short-term. Even though substrate proteins would be phosphorylated

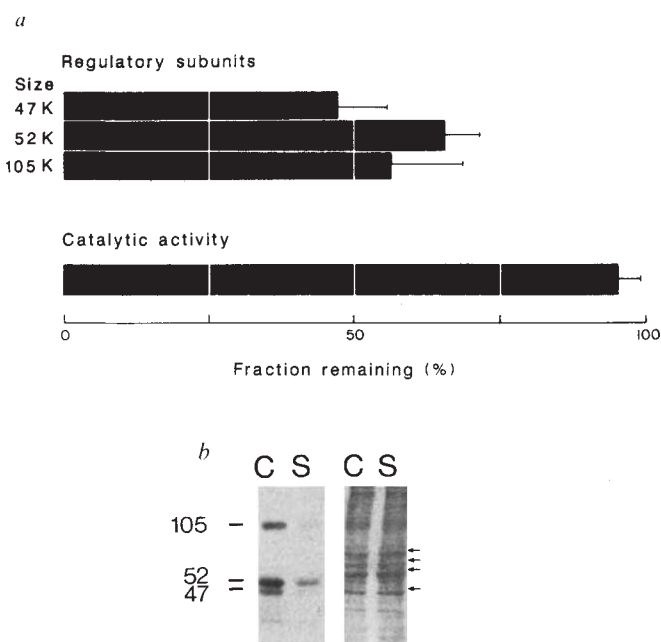


Fig. 3 The effect of treatment with CPT-cAMP on cAMP-dependent protein kinase in synaptosomes. Synaptosomes were prepared according to G. J. Chin, E. Shapiro and J. H. Schwartz (personal communication), resuspended in normal sea water (0.1 ml) either with or without 1 mM CPT-cAMP, and incubated for 2 h at 15 °C. The synaptosomes were subsequently sedimented and then resuspended in Ca^{2+} -free sea water (0.1 ml) for 20 min at 15 °C. This wash was repeated four times to remove CPT-cAMP. The pellets were photolabelled and quantitated as in Fig. 1. *a*, Data from five experiments. Each class of R subunit from CPT-cAMP-treated synaptosomes is significantly (for the M_r 47K and 52K species, $P<0.01$; for the M_r 105K species, $P<0.05$, two-tailed Student's *t*-test) decreased from control (set at 100%). Protein kinase activity was measured as described in Table 1 legend. *b*, Autoradiograph and Coomassie-blue-stained gel from a typical experiment in which synaptosomes were incubated without (lanes C) or with (lanes S) CPT-cAMP. Arrows (from bottom to top) position of *Aplysia* actin (M_r 43K), tubulin (55K), and neurofilament proteins (60K and 65K). Four kinds of control experiments indicate that the diminished photoaffinity labelling of R observed in synaptosomes treated with CPT-cAMP did not result from retention of the analogue in the samples and subsequent interference with the photoaffinity reagent. (1) A similar loss of labelled R occurred if [^{32}P]8- N_3 cAMP was used both to treat synaptosomes made permeable by exposure to 0.2% saponin in 0.1 M Tris-HCl (pH 8.0) containing 5 mM EDTA and EGTA and to photolabel the residual R subunits. (2) Little loss was observed in synaptosomes after a 30-min exposure to CPT-cAMP (the earliest time point assayed). (3) Purified bovine heart R_{II} (refs 27 and 28), added as internal standard to extracts of synaptosomes treated for 2 h, was labelled as heavily as R_{II} added to extracts of untreated synaptosomes. (4) No difference was observed in the extent of photoaffinity labelling of R in extracts of *Aplysia* nervous tissue isolated by gel filtration after treatment with CPT-cAMP (see Fig. 1, legend), indicating that binding of CPT-cAMP (like cAMP) is reversible.

to a greater extent 24 h after training, a facilitatory input to a trained sensory neuron could still enhance cAMP synthesis from the basal (B) to the stimulated (S) concentration. As a result, sensory cells could still show short-term sensitization in addition to long-term sensitization³ by activating the remaining holoenzyme (Fig. 2*b*, pathway 3).

Synthesis of new protein is required for the stable acquisition of synaptic facilitation in *Aplysia*³ and has been implicated in molecular models for long-term memory¹. Newly synthesized proteins, however, require some time to reach synaptic terminals, which can be distant from the cell body. To determine whether

synthesis of new protein is needed for the change in R subunits, *Aplysia* synaptosomes, which incorporate <5% the amount of ^{35}S -methionine compared to an equivalent mass of intact ganglion, were treated with 8-(4-chlorophenylthio)-cAMP (CPT-cAMP), a permeant analogue effective in *Aplysia*¹⁴. R subunits were found to decrease with the time of exposure: after 2 h, a period chosen to approximate the time cAMP is elevated by 90 min of training, R decreased by 40% (Fig. 3); after 24 h, by 72%. Treatment with CPT-cAMP affected neither the amount of catalytic activity (Fig. 3a) nor any of the major proteins stained with Coomassie Blue (Fig. 3b). Much evidence from other species indicates that dissociated R subunits are prone to proteolytic cleavage¹⁵⁻¹⁷. Increased degradation rather than decreased synthesis is thus the most likely explanation for the loss of R in synaptosomes treated with CPT-cAMP.

The requirement that protein must be synthesized in order to establish long-term memory, inferred from experiments³ with dissociated neurons in culture, is not necessarily inconsistent with the post-translational process that we describe. If, as behavioural and neurophysiological studies suggest (refs 3, 6 and 13 and N. Dale *et al.*, personal communication), short-term and long-term sensitization are continuous, there must be an interim molecular mechanism for enhancing release of transmitter before new gene products arrive from the cell body⁴. The cAMP-dependent degradation of R subunits at synapses is a molecular process that could underlie sensitization long after cAMP has been dissipated. The experiments described here do not address the question of how the change in R subunits observed in the intact animal 24 h after behavioural training might depend on the formation of new protein. The duration of the memory might be sustained in the long term if the altered subunit ratio becomes dependent on newly induced gene products arriving at the nerve terminals. If new protein is required to maintain the R-to-C ratio at a diminished value, that protein might be a specific protease¹⁸, an altered version of R more susceptible to cleavage by existing proteases¹⁹ or a DNA-binding protein¹ that directly signals the decrease in gene expression of R.

We thank Drs Gregg Clark, Nicholas Dale, Irving Kupfermann and Eric Kandel for comments on the manuscript. S.M.G. is a trainee of the Public Health Service Medical Scientist Training Program.

Received 9 April; accepted 10 July 1987.

- Goelet, P., Castellucci, V. F., Schacher, S. & Kandel, E. R. *Nature* **322**, 419-422 (1986).
- Kandel, E. R. & Schwartz, J. H. *Science* **218**, 433-443 (1982).
- Montarolo, P. G. *et al. Science* **234**, 1249-1254 (1986).
- Schwartz, J. H. & Greenberg, S. M. *A. Rev. Neurosci.* **10**, 459-476 (1987).
- Bernier, L., Castellucci, V. F., Kandel, E. R. & Schwartz, J. H. *J. Neurosci.* **2**, 1682-1691 (1982).
- Frost, W. N., Castellucci, V. F., Hawkins, R. D. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8266-8269 (1985).
- Eppler, C. M., Palazzolo, M. J. & Schwartz, J. H. *J. Neurosci.* **2**, 1692-1704 (1982).
- Eppler, C. M., Bayley, H., Greenberg, S. M. & Schwartz, J. H. *J. Cell Biol.* **102**, 320-321 (1986).
- Walters, E. T., Byrne, J. H., Carew, T. J. & Kandel, E. R. *J. Neurophysiol.* **50**, 1543-1559 (1983).
- Greenberg, S. M., Bernier, L. & Schwartz, J. H. *J. Neurosci.* **7**, 291-301 (1987).
- Castellucci, V. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7492-7496 (1980).
- Castellucci, V. F., Nairn, A., Greengard, P., Schwartz, J. H. & Kandel, E. R. *J. Neurosci.* **2**, 1673-1681 (1982).
- Scholz, K. P. & Byrne, J. H. *Science* **235**, 685-687 (1987).
- Weiss, K. R., Mandelbaum, D. E., Schonberg, M. & Kupfermann, I. *J. Neurophysiol.* **42**, 791-803 (1979).
- Potter, R. L. & Taylor, S. S. *J. Biol. Chem.* **255**, 9706-9712 (1980).
- Steinberg, R. A. & Agard, D. A. *J. Biol. Chem.* **256**, 10731-10734 (1981).
- Beer, D. G., Butley, M. S. & Malkinson, A. M. *Arch. Biochem. Biophys.* **228**, 207-219 (1984).
- DeLotto, R. & Spierer, P. *Nature* **323**, 688-692 (1986).
- Rogers, S., Willis, R. & Rechsteiner, M. *Science* **234**, 364-368 (1986).
- Corbin, J. D., Sugden, P. H., West, L., Flockhart, D. A., Lincoln, T. M. & McCarthy, D. *J. Biol. Chem.* **253**, 3997-4003 (1978).
- Rangel-Aldao, R., Kupiec, J. W. & Rosen, O. M. *J. Biol. Chem.* **254**, 2499-2508 (1979).
- Kerlavage, A. R. & Taylor, S. S. *J. Biol. Chem.* **257**, 1749-1754 (1982).
- Walter, U., Costa, M. R. C., Breakfield, X. O. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3251-3255 (1979).
- Bubis, J. & Taylor, S. S. *Biochemistry* **24**, 2163-2170 (1985).
- Schaffner, W. & Weissman, C. *Analyt. Biochem.* **56**, 502-514 (1973).
- Schwechheimer, K. & Hofmann, F. *J. Biol. Chem.* **252**, 7690-7696 (1977).
- Rannels, S. R., Beasley, A. & Corbin, J. D. *Meth. Enzym.* **99**, 55-62 (1983).
- Builder, S. E., Beavo, J. A. & Krebs, E. G. *J. Biochem.* **255**, 3514-3519 (1980).

Amelioration of cholinergic neuron atrophy and spatial memory impairment in aged rats by nerve growth factor

W. Fischer, K. Wictorin, A. Björklund, L. R. Williams*, S. Varon† & F. H. Gage*

Department of Histology, University of Lund, S-223 62 Lund, Sweden

Departments of * Neurosciences and † Biology, School of Medicine, University of California, San Diego, La Jolla, California 92093, USA

In aged rodents, impairments in learning and memory have been associated with an age-dependent decline in forebrain of cholinergic function¹, and recent evidence indicates that the cholinergic neurons in the nucleus basalis magnocellularis, the septal-diagonal band area and the striatum undergo age-dependent atrophy²⁻⁵. Thus, as in Alzheimer-type dementia in man, degenerative changes in the forebrain cholinergic system may contribute to age-related cognitive impairments in rodents. The cause of these degenerative changes is not known. Recent studies have shown that the central cholinergic neurons in the septal-diagonal band area, nucleus basalis and striatum are sensitive to the neurotrophic protein nerve growth factor (NGF)⁶⁻¹⁰. In particular, intraventricular injections or infusions of NGF in young adult rats have been shown to prevent retrograde neuronal cell death¹¹⁻¹³ and promote behavioural recovery after damage to the septo-hippocampal connections¹⁴. It is so far not known, however, whether the atrophic cholinergic neurons in aged animals are responsive to NGF treatment. We report here that continuous intracerebral infusion of NGF over a period of four weeks can partly reverse the cholinergic cell body atrophy and improve retention of a spatial memory task in behaviourally impaired aged rats.

Behaviourally impaired ($n=24$) and behaviourally non-impaired ($n=8$) aged rats were selected by previously established criteria^{15,16} on the basis of their performance in a pre-test in the Morris' water maze task¹⁷, from a group of 60 23-25-month-old female Sprague-Dawley rats (retired breeders from ALAB, Stockholm).

Two months after the pre-test, the aged impaired rats were randomly allocated to an experimental ($n=11$) and a control ($n=13$) group. Cannulae attached to modified Alzet Model 2002 mini-osmotic pumps were implanted into the lateral ventricle in all experimental rats, and in half the control rats, as described in detail previously^{12,18}. In the experimental rats the pump was filled with artificial cerebrospinal fluid (CSF) containing $22\text{ }\mu\text{g ml}^{-1}$ of 7S NGF (equivalent to about $4,000\text{ BU ml}^{-1}$, where 1 BU/ml = minimal concentration for maximal response *in vitro*) and 0.1 mg ml^{-1} rat serum albumin (Sigma). In the control rats the pump was filled with artificial CSF (with rat serum albumin). The estimated NGF dose was $1\text{ }\mu\text{g}$ or about 200 BU per week. The pumps were re-filled once (after 15 days) and the rats were killed for acetylcholinesterase (AChE) histochemistry (Fig. 2) after 28 days of continuous infusion.

The rats were retested on the Morris' water maze task twice during the NGF infusion period: between day 8 and day 12 (test week 1) and between day 24 and day 27 (test week 2) (Fig. 1). The performance of the aged impaired rats remained impaired compared to the aged but non-impaired rats during the first test week ($P<0.005$), and the NGF-treated animals did not differ from the control rats. In test week 2, the NGF-treated rats were significantly improved in their overall performance compared to the aged control rats ($P<0.025$ for both escape latency and distance), and their performance was no longer different from that of the aged non-impaired group. The improvement between the first and second test weeks in the NGF-treated rats ($P<0.01$ for both escape latency and distance)

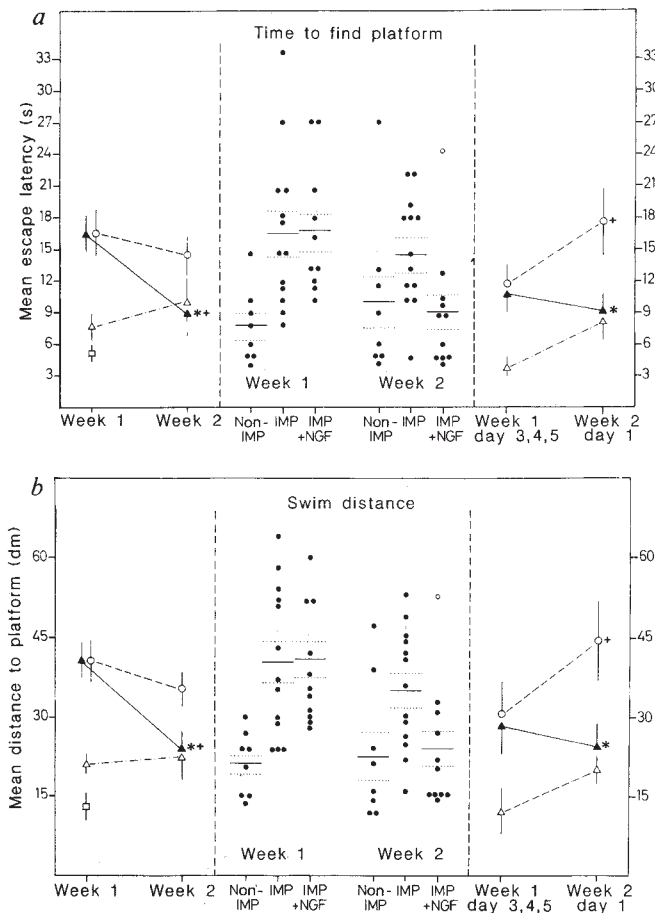


Fig. 1 Water maze test. Performance of three groups of aged rats (Δ , aged non-impaired; $\circ$, aged impaired; $\blacktriangle$, NGF-treated aged impaired) in test weeks 1 and 2, expressed in *a* as the time to reach the hidden platform (escape latency), and in *b* as the swim distance to reach the platform. A group of identically tested young (three-month-old) rats ($\square$) has been included for comparison. The rats were given a total of 36 trials over 5 days in week 1, and 28 trials over 4 days in week 2, as described previously^{15,16}. Left, mean performance ($\pm$ s.e.m.) over all trials. Middle, individual mean escape latency or swim distance scores for the rats in each group, with the means (solid bars) and s.e.m. (pairs of broken bars) of each group indicated; open circles, performance of a rat in which the pump was found to be disconnected from the infusion cannula when it was killed (IMP, impaired; NGF, NGF treated). Right, memory retention expressed as the difference between the mean performance in the last three days of test week 1 and the first test day of test week 2.

Methods. In the statistical analysis for test week 1, one-way ANOVA showed differences between groups both for the whole test week (latency: $F=6.20$, $P<0.01$; distance: $F=7.87$, $P<0.005$) and for days 3–5 (latency: $F=3.98$, $P<0.05$; distance: $F=5.10$, $P<0.025$). Both aged impaired groups differed from the non-impaired group ($P<0.01$; Student's *t*-test) but not from each other. For test week 2, one-way ANOVA showed differences between groups for day 1 (latency: $F=9.70$, $P<0.005$; distance: $F=4.80$, $P<0.025$) and on the distance measured for the whole week ($F=3.36$, $P<0.05$). The NGF-treated group differed from the aged impaired controls both on day 1 (latency: $P<0.025$; distance: $P<0.05$; asterisks in right panels) and on the means over the whole second week (latency and distance: $P<0.025$; asterisks in left panels). The NGF-treated group alone showed significantly improved performance in test week 2 compared to test week 1 (crosses in left panels). In the retention test (right panels) the age-impaired controls, but not the NGF-treated aged-impaired rats, showed significantly impaired performance on the first day of week 2 compared to the performance on days 3–5 of week 1 (latency: $P<0.025$; distance: $P<0.05$; crosses in right panels). Open circle, rat found to have pump disconnected from infusion in cannula when it was killed. This animal was included in behavioural analyses.

was primarily due to improved retention of the previously acquired performance over the 12 days separating the two tests. As shown in Fig. 1 (right hand panels) the NGF treated rats performed as well on the first day of the second test as they did at the end of the previous test, whereas the control aged impaired rats showed a significant retention deficit. During the second test week the aged impaired controls reacquired their performance and, thus, the difference between the two groups was no longer significant by the end of the week. This indicates that the main NGF effect was due to improved retention of the acquired performance over the 12 days which separated the two tests, whereas acquisition (the rats' ability to learn the task) was unaffected.

By the end of the second test week the aged impaired rats and a separate group of young, three-month-old rats were treated with an irreversible AChE inhibitor¹⁹ and processed for AChE histochemistry (Fig. 2). Twelve of the brains which had sufficiently high quality of AChE-positive cell body visualization

were taken for image analysis using a Zeiss-Kontron IBAS II instrument (Fig. 2).

In all four regions analysed (medial septum, MS; the vertical limb of the diagonal band, VDB; striatum; nucleus basalis magnocellularis, NBM), the size of the AChE-positive cell bodies was significantly reduced in the aged impaired control rats (compare squares and circles in Fig. 2, and Fig. 3*a* and *c*). The reduction amounted to $\sim 25\%$ in MS and to $\sim 40\%$ in striatum, VDB and NBM, and there were no significant differences between the two sides. In the NGF-treated rats the AChE-positive cell bodies were significantly larger on the side of the infusion (the right side) in the NBM in all five rats (Fig. 3*d*), in the striatum in four of the five rats, and in the VDB in two of the five rats (asterisks in Fig. 2). The overall increase in cross-sectional area of the AChE-positive cell bodies was 18.8% in the NBM and 21.0% in the striatum (Table 1).

The number of AChE-positive cell bodies per unit area did not differ between the NGF-treated and the vehicle-treated aged

Table 1 AChE-positive cell body size and spatial memory retention for young rats and treated and untreated aged rats

		MS	VDB	Striatum	NBM
Aged impaired NGF-treated	L	129.8 $\pm$ 9.5	136.6 $\pm$ 7.9	122.0 $\pm$ 10.7	155.6 $\pm$ 6.2
(<i>n</i> = 5)	R	136.8 $\pm$ 8.8	143.4 $\pm$ 5.6	156.2 $\pm$ 11.7*	184.8 $\pm$ 10.0*
Aged impaired, vehicle-treated	L	119.9 $\pm$ 5.3	146.4 $\pm$ 6.3	140.1 $\pm$ 6.5	165.4 $\pm$ 3.9
(<i>n</i> = 4)	R	132.0 $\pm$ 7.4	133.9 $\pm$ 10.0	143.4 $\pm$ 4.8	162.1 $\pm$ 3.8
Young	L	170.2 $\pm$ 7.4	226.6 $\pm$ 11.2	236.8 $\pm$ 6.6	268.8 $\pm$ 23.4
(<i>n</i> = 3)	R	172.7 $\pm$ 2.6	229.5 $\pm$ 16.7	233.3 $\pm$ 3.8	270.1 $\pm$ 18.9

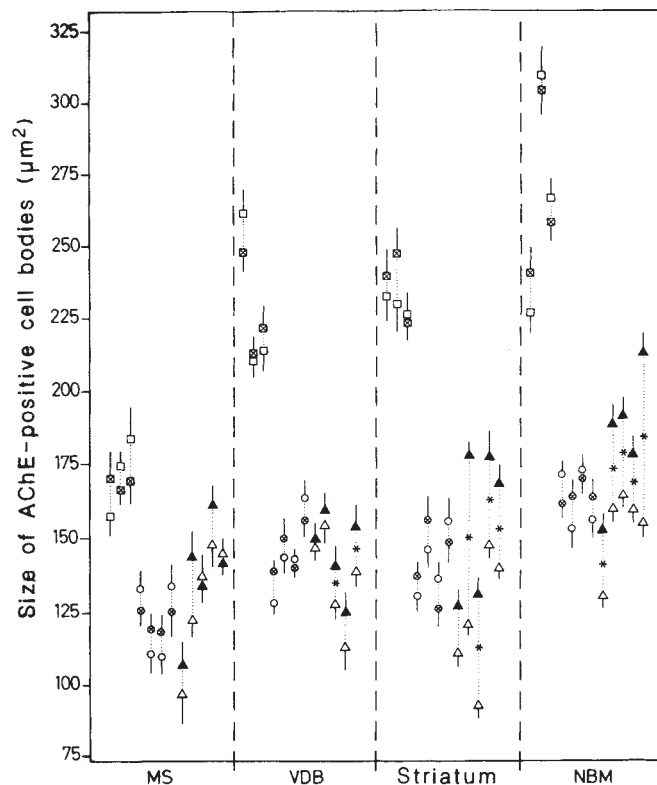
Values are means $\pm$ s.e.m. of AChE-positive cell body size given as cross-sectional area in μm^2 . The mean escape latency ($\pm$ s.e.m.) in the memory retention test on day 1 of the second test week is also given for the two subgroups of aged impaired rats which were used in the image analysis.

* $P<0.01$ compared to the non-infused (left) side; Student's related *t*-test.

† $P<0.01$ compared to vehicle injected controls; Student's unrelated *t*-test.

Fig. 2 Size of AChE-positive cell bodies, expressed as cross-sectional area in μm^2 (MS, medial septum; VDB, vertical limb of the diagonal band; striatum, head of caudatus-putamen nucleus; NBM, nc. basal magnocellular nucleus). Symbols designate young rats ($\square$, left; $\boxtimes$, right); and aged impaired rats infused with vehicle ($\circ$, left; $\otimes$, right) or NGF ($\triangle$, left; $\blacktriangle$, right). Each pair of symbols, connected by a dotted line, represents one animal.

Methods. The rats were treated with 2 mg per kg (intramuscularly, i.m.) of diisopropylfluorophosphate (DFP) (Sigma) and 4.5 h later were perfused with ice-cold 4% phosphate-buffered paraformaldehyde. The brains were serially sectioned in the coronal plane at 20 μm on a cryostat. Every third section was stained for AChE histochemistry²⁸ using 10^{-4} M ethopropazine (profenamine) to inhibit non-specific esterases. Alternating sections were stained with cresyl violet. For image analysis, every second AChE-stained section throughout the extent of the septal-diagonal band area and the basal nucleus region, and every third section through the head of the caudate-putamen, were used for quantitation of size and number of AChE-positive cell bodies. The delineation of medial septum (MS) and the vertical limb of the diagonal band (VDB) followed previously defined borderlines²⁹. The area included in the striatum measurement covered the entire cross-section of the head of the caudate-putamen down to the level of the anterior commissure, starting rostrally at the level of fusion of the corpus callosum and ending caudally at the level of the crossing of the anterior commissure. The basal nucleus (NBM) was defined as the AChE-positive neurons located in the globus pallidus and the internal capsule. They were measured from the level of the caudal septum, rostrally, to the level of the tail of the caudate-putamen. The number of measured sections were about 15 for MS, 15 for VDB, 15 for striatum and about 20 for NBM. The mean number of cells measured on each side of each brain was 2,270 in the young rats and 1,130 in the aged ones. Cells were defined (under $\times 4$ objective magnification) as AChE-positive bodies with a diameter of $>8 \mu\text{m}$ along the minor axis and a cross-sectional area $>70 \mu\text{m}^2$. These criteria were used in order not to exclude any possible population of small atrophied AChE-positive neurons in the aged rats.



rats. However, as will be reported elsewhere (W.F. *et al.*, in preparation), the cell numbers were markedly reduced in the aged impaired rats compared to the young rats ($-62 \pm 4.6\%$ in MS; $-60 \pm 6.3\%$ in VDB; $-34 \pm 10.3\%$ in striatum and $-42 \pm 1.1\%$ in NBM) (compare Fig. 3a and b with c and d).

Our results are consistent with previous *in vivo* observations on the effects of intracerebrally administered NGF on biochemical and behavioural recovery after partial lesions of the septo-hippocampal cholinergic pathway in young adult rats^{7,14}. The ability of NGF to counteract retrograde degeneration of the septo-hippocampal cholinergic neurons after axotomy suggests that these effects may be due to a trophic, protective action on the partially damaged cholinergic neurons.

The significant increases in size of the AChE-positive neurons in NBM and striatum on the side ipsilateral to the NGF infusion indicate that NGF can exert significant trophic effects also on the atrophic cholinergic forebrain system in aged, but otherwise intact, rats. Because these effects were seen only on the side ipsilateral to the infusion, it seems likely that the infused NGF exerted its actions primarily through local diffusion into the brain tissue, including the part of the striatum close to the infusion site. This may allow the infused NGF to be taken up by the striatal cholinergic interneurons and by the axons of the basalo-cortical cholinergic neurons (some of which are known to pass through the striatum²⁰).

The performance of young rats in the water-maze task is sensitive to drugs or lesions affecting the forebrain cholinergic system²¹⁻²² and there are abundant data in support of the idea that impaired function in the forebrain cholinergic projection system contributes to the decline in learning and memory that is seen in ageing rodents (see refs 1 and 23 for review). Therefore, it seems possible (although by no means proven) that the improved spatial memory with NGF infusion in the behaviourally impaired aged rats could be due, at least in part, to the effect of NGF on the forebrain cholinergic neurons. It is notable that the NGF effect seemed to be specifically on one

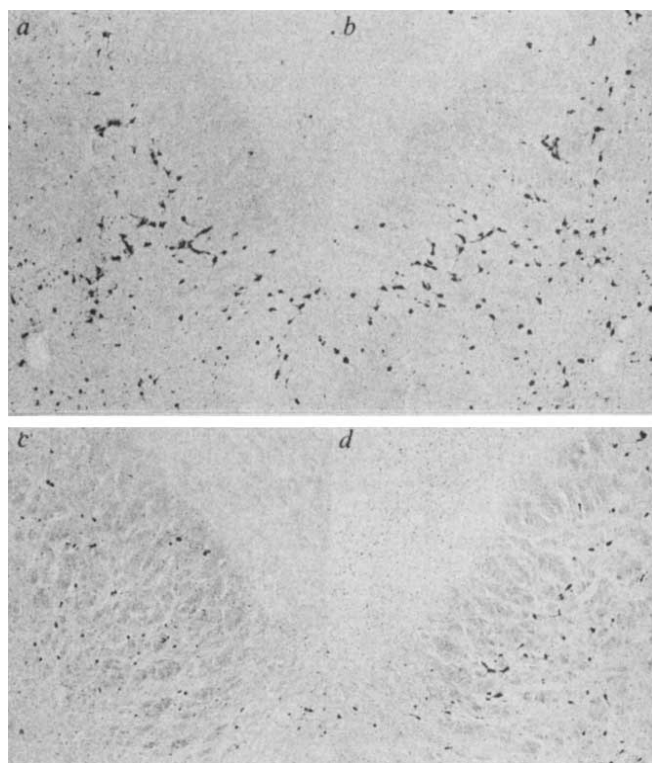


Fig. 3 The basal nucleus at the level of the caudal globus pallidus on the left and right side of a young control rat (a and b) and on the left (non-infused) side (c) and the right (infused) side (d) in an NGF-treated aged impaired rat. Note that the AChE-positive cells on the NGF-infused side (d) are larger than the ones on the non-infused side (c), and that the cells in the aged rat are smaller and fewer in number than in the young control rat.

particular aspect, namely spatial memory retention between the two tests, whereas parameters of acquisition in the two tests were unaffected.

Recent studies^{11-13,24-27} have suggested that target-derived trophic factors are important for the maintenance of the basal forebrain cholinergic projection system, and that NGF may be important in this trophic regulation. The present results raise the possibility that the age-dependent structural and functional deterioration of the forebrain cholinergic neurons, implicated in Alzheimer's disease, may be due to a progressive failure of this retrograde trophic support system, and that exogenous supply of NGF may be able, at least partly, to substitute for this age-dependent defect.

We thank Bente Falck, Alicja Flasch, Catarina Friberg, Ulla Jarl and Yvette Jönsson for technical assistance. The study was supported by grants from the National Institutes of Health and Aging, from the Swedish MRC, the office of Naval Research, the Margret and Herbert Hoover Jr Foundation and the Greta and Johan Kock Foundation.

Received 27 May; accepted 22 June 1987.

1. Bartus, R. T., Dean, R. L., Beer, B. & Lippa, A. S. *Science* **217**, 408-416 (1982).
2. Hornberger, J. C., Buell, S. J., Flood, D. G., McNeill, T. H. & Coleman, P. D. *Neurobiol. Aging* **6**, 269-275 (1985).

3. Bieganski, A., Greenberger, V. & Segal, M. *Neurobiol. Aging* **7**, 215-217 (1986).
4. Luine, V. N., Renner, K. J., Heady, S. & Jones, K. J. *Neurobiol. Aging* **7**, 193-198 (1986).
5. Albanese, A., Altavista, M. C., Rossi, P., Iacopino, C. & Gozzo, S. *Neurosci. Lett.* **26**, S88 (1986).
6. Schwab, M. E., Otten, U., Agid, Y. & Thoenen, H. *Brain Res.* **168**, 473-483 (1979).
7. Gnahn, H., Hefti, F., Heumann, R., Schwab, M. E. & Thoenen, H. *Devl Brain Res.* **9**, 45-52 (1983).
8. Hefti, F., Hartikka, F., Eckenstein, H., Gnahn, R., Heumann, R. & Schwab, M. *Neuroscience* **14**, 55-68 (1985).
9. Mobley, W. C., Rutkowski, J. L., Tennekoon, G. I., Buchana, K. & Johnston, M. V. *Science* **229**, 284-286 (1985).
10. Richardson, P. M., Verge, V. M. K. & Riopelle, R. J. *J. Neurosci.* **6**, 2312-2321 (1986).
11. Hefti, F. *J. Neurosci.* **6**, 2155-2162 (1986).
12. Williams, L. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 9231-9235 (1987).
13. Kromer, L. F. *Science* **235**, 214-216 (1987).
14. Will, B. & Hefti, F. *Behav. Brain Res.* **17**, 17-24 (1985).
15. Gage, F. H., Björklund, A., Stenevi, U., Dunnett, S. B. & Kelly, P. A. T. *Science* **225**, 533-536 (1984).
16. Gage, F. H., Kelly, P. A. T. & Björklund, A. *J. Neurosci.* **4**, 2856-2866 (1984).
17. Morris, R. G. M. *J. Neurosci. Meth.* **11**, 47-60 (1984).
18. Williams, L. R. *et al. Expl Neurol.* **95**, 743-754 (1987).
19. Butcher, L. L. in *Handbook of Chemical Neuroanatomy* Vol. 1 (eds Björklund, A. & Hökfelt, T.) 1-49 (Elsevier, Amsterdam, 1983).
20. Woolf, N. J., Hearn, M. C. & Butcher, L. L. *Neurosci. Lett.* **66**, 281-286 (1986).
21. Sutherland, R. J., Whishaw, I. Q. & Regehr, J. C. *J. Comp. physiol. Psychol.* **96**, 563-573 (1982).
22. Wishaw, I. Q., O'Connor, W. T. & Dunnett, S. B. *Behav. Brain Res.* **17**, 103-115 (1985).
23. Gage, F. H. & Björklund, A. *A. Rev. Physiol.* **48**, 447-459 (1986).
24. Appel, S. H. *Ann. Neurol.* **10**, 499-505 (1981).
25. Hefti, F. & Weiner, W. J. *Ann. Neurol.* **20**, 275-281 (1986).
26. Sofroniew, M. V., Pearson, R. C. A., Eckenstein, F., Cuellar, A. C. & Powell, T. P. S. *Brain Res.* **289**, 370-374 (1983).
27. Sofroniew, M. V., Isacson, O. & Björklund, A. *Brain Res.* **378**, 409-415 (1986).
28. Geneser-Jensen, F. A. & Blackstad, T. W. Z. *Zellforsch. mikrosk. Anat.* **144**, 693-720 (1971).
29. Gage, F. H., Wictorin, K., Fischer, W., Williams, L. R., Varon, S. & Björklund, A. *Neuroscience* **19**, 241-256 (1986).

The expression of hybrid HIV:Ty virus-like particles in yeast

Sally E. Adams*, Keith M. Dawson*, Keith Gull†, Susan M. Kingsman† & Alan J. Kingsman*†§

* Department of Molecular Biology, British Bio-technology Ltd, Brook House, Watlington Road, Cowley, Oxford OX4 5LY, UK
† Department of Biochemistry, University of Oxford, South Parks Road, Oxford, UK
‡ Biological Laboratory, University of Kent, Canterbury, UK

The yeast retrotransposon, Ty, encodes a set of proteins that are assembled into virus-like particles, Ty-VLPs (refs 1, 2). These proteins include Ty-VLP structural proteins, a protease that mediates cleavage of primary translation products and a reverse transcriptase^{1,3}. The major structural components of Ty-VLPs are proteolytic products of the primary translation product, p1 (ref. 3). We have recently shown that protein p1 alone can form Ty-VLPs (ref. 3). Here we demonstrate that p1 fusion proteins, comprising most of p1 and part of human immunodeficiency virus (HIV) protein gp120, form hybrid HIV:Ty-VLPs. These hybrid particles provide a rapid means of preparing and evaluating HIV antigens for a variety of immunological purposes.

The genetic organization of the Ty element is shown in Fig. 1a. It is 5.9 kilobases (kb) in length, with 340 base pair (bp) long terminal repeats called delta sequences^{4,5}. A full-length 5.7-kb transcript, which starts in the left delta and ends in the right^{6,7}, contains two overlapping genes, TYA and TYB (refs 8-11). TYA is expressed by simple translation to produce protein p1 (refs 8, 12, 13). TYB, on the other hand, is expressed as a TYA:TYB fusion protein, p3, via a ribosomal frameshift within the overlap region^{8,10,13,14}. It is likely, though unproven, that p1 and p3 assemble into a pre-Ty-VLP and then undergo proteolytic cleavage to produce a mature Ty-VLP containing functional products³ (Fig. 1a). The mature Ty-VLPs contain the components required for transposition via an RNA

intermediate^{1-3,15}, and we have suggested that this is their function *in vivo*^{1,3}.

The observation that protein p1, the primary translation product of the TYA gene, can produce Ty-VLPs in the absence of proteolytic cleavage³ suggested that it may be possible to construct Ty-VLPs from p1 fusion proteins and thereby create particulate, polyvalent derivatives of potentially any protein or protein domain. An expression vector, pMA5620, was constructed for this purpose (Fig. 1b). This is an *E. coli*/yeast shuttle vector with the promoter from the yeast PGK gene^{16,17} driving expression of the first 381 codons of TYA. At codon 381 is a BamHI linker (CCGGATCCGG) which is followed by termination codons in all three reading phases and a transcription terminator. Any coding sequence or part of a coding sequence can be inserted at the unique BamHI site in pMA5620 to produce a TYA fusion gene.

To produce TYA:HIV fusions, two HIV env fragments were inserted, in frame, into the BamHI site of pMA5620. Fragment hiv5 is a *Dra*I:*Dra*I fragment, comprising codons 130 to 342 of env, and fragment hiv8 is a *Sau*3a:*Sau*3a fragment, comprising codons 274 to 466 (refs 18, 19). The resulting plasmids are designated pMA5620-hiv5 and pMA5620-hiv8 and they have the general structure shown in Fig. 1b. Yeast transformants containing these plasmids are designated T5620-hiv5 and T5620-hiv8, respectively. Both hiv5 and hiv8 come from the region of env encoding the viral surface protein gp120 (refs 18, 19) and they overlap by 68 codons (Fig. 1c).

When extracts of transformants containing pMA5620 (T5620) are fractionated on 15-45% sucrose gradients³, particulate Ty-VLP proteins can be found towards the lower half of the gradient. These are readily identified as Ty-VLP proteins by using anti-Ty-VLP antiserum in a Western blot of sucrose gradient fractions run out on SDS polyacrylamide gels (Fig. 4a). No such particulate proteins are seen in extracts of the same yeast strain containing a plasmid (pMA91) (refs 16, 17) which is identical to pMA5620 but lacks the TYA component. The particulate nature of the protein produced by pMA5620 was confirmed by examining peak fractions from the gradient in the electron microscope. Figure 2a shows examples of Ty-VLPs seen in these fractions. They are very similar to 'wild type' particles described previously^{1,3}, having a dark core and pale outer 'shell'. This shows that the first 381 amino acids from the 439 amino acids of the p1 protein are sufficient for Ty-VLP formation.

§ To whom correspondence should be addressed.

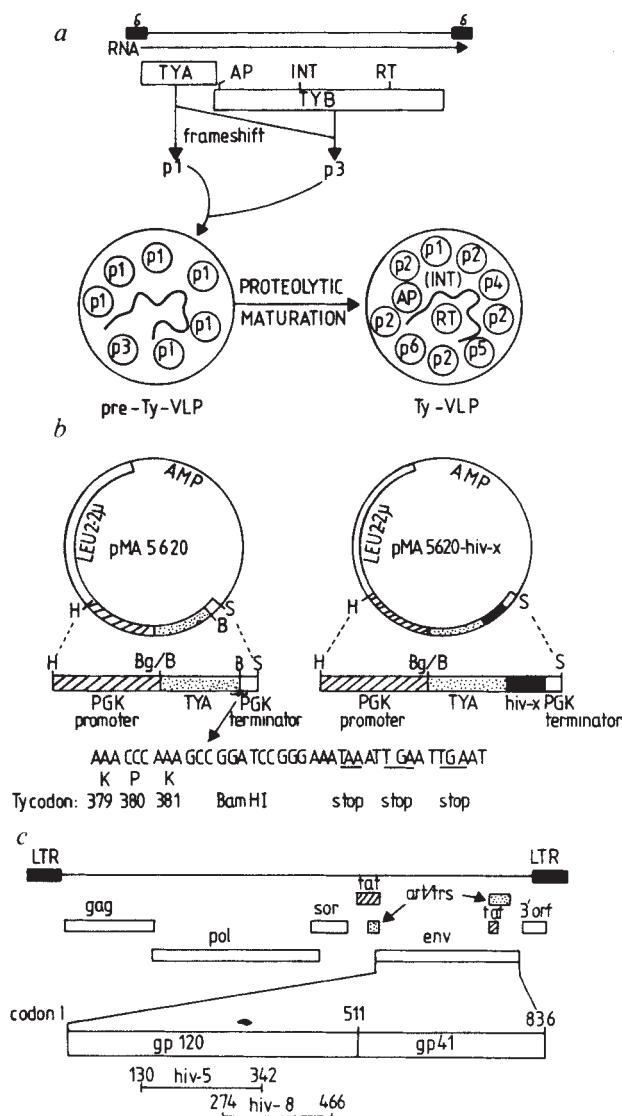


Fig. 1 *a*, Genetic organization and protein products of the yeast Ty element. The top line shows the 5.9-kb element, with the long terminal repeats and delta sequences marked as closed boxes. The major 5.7-kb Ty RNA species is shown^{6,7}. TYA and TYB are shown as open boxes, with regions of TYB showing amino-acid homology to retroviral *pol* genes marked as: AP for acid protease active site²³; INT for integrase^{10,11}; RT for reverse transcriptase^{10,11}. Proteins p1 and p3 (refs 12, 13) are shown and the components of the pre-Ty-VLP and Ty-VLP are p1 and its derivatives p2, p4, p5 and p6 (ref. 3); p3 and its derivatives reverse transcriptase (RT) (refs 1, 3), protease (AP) (ref. 3) and integrase (INT). INT is in parentheses because it has not been detected. The wavy line represents Ty RNA. *b*, The structure of pMA5620 and pMA5620 containing HIV fragments hiv5 or hiv8. Plasmid pMA5620 is a pMA91 (refs 6, 17) derivative containing the first 381 codons of the TYA gene from Ty1-15 (ref. 24). The structure of the sequence surrounding the unique *Bam*HI site is shown. pMA5620-hiv-x is a generalized diagram representing pMA5620 containing any HIV fragment inserted at the *Bam*HI site. Thin line, pBR322; open box, 2 μ :LEU2 selection and replication module²⁵; hatched box, PGK promoter sequences^{16,17}; stippled box, codons 1 to 381 of TYA; closed box, HIV fragment (hiv5 or hiv8). A PGK terminator fragment containing stop codons in all three reading phases and a transcriptional stop signal is marked. *c*, Fragments hiv5 and hiv8 on the HIV genetic map. The genetic organization of HIV is shown with the *env* gene expanded^{18,19}. The coding sequences for the two *env* products, gp120 and gp41, are marked. The coordinates given for hiv5 and hiv8 are codons from the start of the *env* open reading frame.

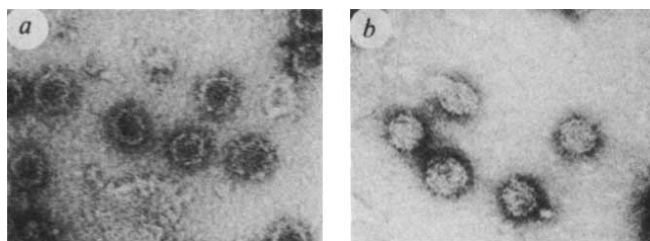


Fig. 2 Electron micrograph of *a*, T5620 and *b*, 5620-hiv5 particles. Both types of particle are about 60 nm in diameter and were prepared for electron microscopy as previously described¹.

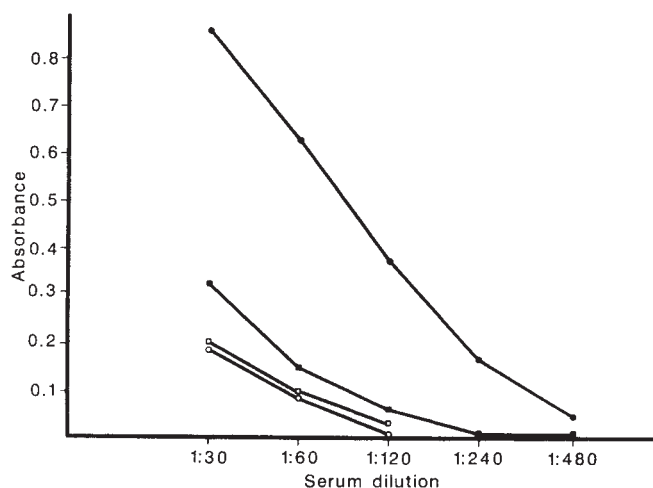


Fig. 3 Serum titration of anti HIV:Ty-VLP antisera reactive against HIV1. The reactivities of rabbit antisera raised against T5620-hiv5-VLPs (●), the corresponding pre-immune sera (○); antisera raised against T5620-hiv8-VLPs (■) and the corresponding pre-immune sera (□) with HIV1 antigen were assayed using an indirect ELISA test.

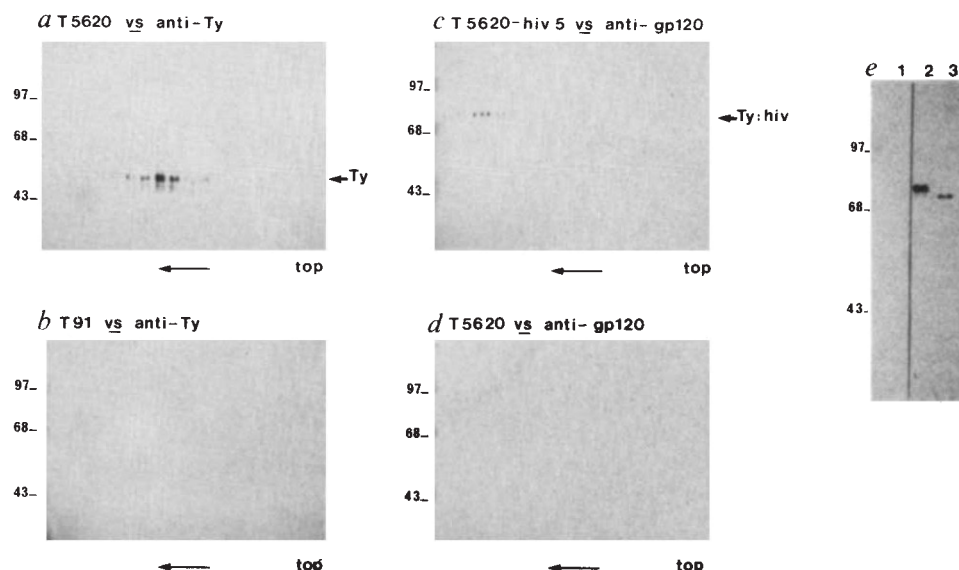
Methods. Rabbits were immunized with three injections of 200 μ g protein of either T5620-hiv5-VLPs or T5620-hiv8-VLPs using standard procedures³. Reactivity of the antisera against HIV1 was assayed using an ELISA test kit (Wellcozyme VK50, Wellcome Diagnostics). Antibody bound to immobilized HIV1 antigen was detected using a biotinylated anti-rabbit second antibody followed by a streptavidin bridge, biotinylated horseradish peroxidase (Amersham) and tetramethylbenzidine.

Similarly, T5620-hiv5 extracts contain a particulate protein of relative molecular mass (M_r) of ~73,000 (73K), the size expected for the fusion protein produced by pMA5620-hiv5. In this case, the protein is identified using an anti-gp120 monoclonal antibody (Fig. 4c). Similar results are obtained with T5620-hiv8, although the signals resulting from the Western blot are weaker (data not shown). When particles are concentrated and their proteins run on SDS-polyacrylamide gels, however, probing with the anti-gp120 monoclonal antibody gives a clear signal (Fig. 4e). No signal is generated in similar experiments with a T5620 extract (Fig. 4d). When peak fractions from T5620-hiv5 and T5620-hiv8 were examined in the electron microscope, particles with a distinct morphology could be seen. They were the same size as 'wild type' Ty-VLPs but without a distinct core and shell (Fig. 2b). T5620-hiv5 and T5620-hiv8 particles looked identical. This shows that T5620-hiv5 and T5620-hiv8 produce p1: HIV fusion proteins and that these fusion proteins assemble into hybrid HIV:Ty-VLPs.

The ease of construction and purification of hybrid Ty-VLPs makes them ideal as a source of antigen. To test the immunogenicity of T5620-hiv5 and T5620-hiv8 particles, concentrated particle preparations were injected into rabbits. The reactivities of the resulting antisera against HIV1 antigen were then tested in

Fig. 4 Western blot analyses of Ty-VLPs and HIV:Ty-VLPs. Sucrose density gradient fractions from extracts of: *a, d*, T5620; *b*, T91-containing plasmid pMA91 (refs 6, 17); and *c*, T5620-hiv5 were run on SDS-polyacrylamide gels. The proteins were transferred to nitrocellulose and probed either with anti Ty-VLP antibody (*a, b*) or anti-gp120 monoclonal antibody (Dupont NEA-7054) (*c, d*). The positions of Ty-VLPs and hybrid HIV:Ty-VLPs are marked. *e*, Western blot of VLPs purified from extracts of T5620 (lane 1), T5620-hiv5 (lane 2) and T5620-hiv8 (lane 3) and probed with anti-gp120 monoclonal antibody.

Methods. Ty-VLPs were isolated as described previously³. Hybrid HIV:Ty-VLPs were purified by the same procedure. Aliquots (25 μ l) of fractions from 15–45% (w/v) sucrose gradients were electrophoresed on 9% SDS polyacrylamide gels²⁶ and the proteins electroblotted to nitrocellulose²⁷. The filters were probed with rabbit polyclonal antisera raised against Ty-VLPs (ref. 3) or a mouse monoclonal antibody reactive against the HIV1 gp120 protein (Dupont). Bound antibody was detected using goat anti-rabbit or anti-mouse second antibody followed by a peroxidase-antiperoxidase (PAP) complex (Sigma) and diaminobenzidine²⁸. VLPs were concentrated by centrifugation of the peak fractions at 100,000g for 1 h at 4 °C and resuspended overnight in 10 mM Tris, pH 7.4, 10 mM NaCl.



an indirect enzyme-linked immunosorbent assay (ELISA) (Fig. 3). The antisera were designated anti-hiv5 and anti-hiv8. Both antisera gave positive results in the ELISA, although the reaction of anti-hiv8 antisera was much weaker than that of anti-hiv5 antisera. Titres were approximately 1/500 and 1/100 for anti-hiv5 and anti-hiv8, respectively. The weaker reaction obtained with anti-hiv8 antisera was not the result of a generally poor immune response in the rabbit, because the titres for anti-hiv5 and anti-hiv8 antisera were both approximately 1/600 in an ELISA using T5620 Ty-VLPs as antigen. Therefore, whereas the response to the Ty components was the same for both types of particle, the response to the HIV components was different. This presumably reflects a difference in the immunogenicity of the two HIV fragments in these constructions. Taken together, these data show that, at least for the hiv5 fragment, hybrid HIV:Ty-VLPs are an effective source of HIV antigen for the production of antisera.

A key feature of the Ty-VLP system that we describe is the ease with which pure Ty-VLPs can be prepared for use as reagents or to raise antisera. This makes it possible to survey a genome, such as that of HIV, for regions that encode important antigenic determinants that induce neutralizing and protective antibodies. In situations where conformational antigens may be significant, the choice of fragment to encode folding domains may be important^{18,19}. The usefulness of the system is also

illustrated by the reaction of T5620-hiv5 and T5620-hiv8 particles with the Dupont anti-gp120 monoclonal antibody; this result maps the location of the epitope for this antibody to the overlap between fragments hiv5 and hiv8, corresponding to amino acids 274 to 342 of *env*^{20,21}. A strategy using multiple overlapping fragments linked to pMA5620 to produce defined particulate antigens could map any epitope.

The formation of hybrid Ty-VLPs appears to be a robust process. We have been able to produce hybrid particles carrying the whole of an α 2-interferon (20K), the product of the bovine papilloma virus E1 and E2 genes (38K and 43K) and part of an influenza virus haemagglutinin (8K) (S.E.A. and M.H. Malim, unpublished data). These hybrid particles elicit the production of antisera against the non-Ty components.

Another feature of this system is that Ty-VLPs constitute a means of presenting antigen to the immune system in a particulate, polyvalent form. It is thought that such structures may be more effective in inducing immunity than the monomeric proteins produced by conventional recombinant DNA strategies²². Hybrid Ty-VLPs could provide, therefore, a new approach to vaccine production and could eventually contribute to the development of an AIDS vaccine.

We thank Frank Caddick for artwork, John Senior for assistance with immunology and Ray Newsam for assistance with the electron microscopy.

Received 17 June; accepted 4 August 1987.

- Mellor, J. *et al.* *Nature* **318**, 583–586 (1985).
- Garfinkel, D. J., Boeke, J. D. & Fink, G. R. *Cell* **42**, 507–517 (1985).
- Adams, S. E. *et al.* *Cell* **49**, 111–119 (1987).
- Roeder, G. S. & Fink, G. R. in *Mobile Genetic Elements* (ed. Shapiro, J. A.) 299–328 (Academic, New York, 1983).
- Williamson, V. M. *Int. Rev. Cytol.* **83**, 1–15 (1983).
- Elder, R. T., St. John, T. P., Stinchcomb, D. T. & Davis, R. W. *Cold Spring Harb. Symp. quant. Biol.* **45**, 581–583 (1980).
- Elder, R. T., Loh, E. Y. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2432–2436 (1983).
- Mellor, J. *et al.* *Nature* **313**, 243–246 (1985).
- Fulton, A. N. *et al.* *Nucleic Acids Res.* **13**, 4097–4112 (1985).
- Clare, J. & Farabaugh, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2829–2833 (1985).
- Hauber, J., Nelbock-Hochstetter, P. & Feldman, H. *Nucleic Acids Res.* **13**, 2745–2758 (1985).
- Dobson, M. J. *et al.* *EMBO J.* **3**, 1115–1119 (1984).
- Mellor, J. *et al.* *Nucleic Acids Res.* **13**, 6249–6263 (1985).
- Wilson, W., Malim, M. H., Mellor, J., Kingsman, S. M. & Kingsman, A. J. *Nucleic Acids Res.* **14**, 7001–7015 (1986).
- Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. *Cell* **40**, 491–500 (1985).
- Mellor, J. *et al.* *Gene* **24**, 1–14 (1983).
- Kingsman, S. M. & Kingsman, A. J. in *Interferons, SGM Symposium* (eds Burke, D. C. & Morris, A.) 35, 221–254 (Cambridge, University Press, 1983).
- Colman, P. M. *et al.* *Nature* **326**, 358–363 (1987).
- Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. *Science* **233**, 747 (1986).
- Ratner, L. *et al.* *Nature* **313**, 277–284 (1985).
- Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* **40**, 9–17 (1985).
- Valenzuela, P. *et al.* *Biotechnology* **3**, 323–326 (1985).
- Toh, H., Ono, M., Saigo, K. & Miyata, T. *Nature* **315**, 691–692 (1985).
- Kingsman, A. J., Gimlich, R. L., Clarke, L., Chinault, A. C. & Carbon, J. A. *J. molec. Biol.* **145**, 619–632 (1981).
- Beggs, J. D. *Nature* **275**, 104–109 (1978).
- Laemmli, U. K. *Nature* **227**, 68–85 (1970).
- Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).
- de Blas, A. L. & Cherwinski, H. M. *Analyt. Biochem.* **113**, 214–219 (1983).

Depletion of the predominant B-cell population in immunoglobulin μ heavy-chain transgenic mice

Leonore A. Herzenberg*, Alan M. Stall*, Jonathan Braun§, David Weaver‡, David Baltimore‡, Leonard A. Herzenberg* & Rudolf Grosschedl†

* Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

† Department of Microbiology and Immunology, University of California, San Francisco, California 94143, USA

‡ Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA

§ Department of Pathology, University of California, Los Angeles, California 90024, USA

The transgenic mouse line M54 was generated by introducing a functionally-rearranged immunoglobulin μ heavy-chain gene into the germ line of a C57B1/6 inbred mouse¹. Previous examination of the antibodies produced by B-cell hybridomas derived from transgenic M54 mice showed that the presence of the μ transgene grossly altered the immunoglobulin repertoire of unimmunized animals, suggesting that these mice suffer from a serious immunoregulatory perturbation². Studies presented here introduce a new perspective on this functional defect. We show that the lymphoid tissues from these transgenic mice lack virtually all conventional bone-marrow-derived B cells, which constitute the predominant B-cell population in normal mice and which typically produce primary and secondary antibody responses to T-cell-dependent antigens³⁻⁶. Moreover, the bone marrow from transgenic M54 mice is depleted of pre-B lymphocytes, indicating a serious defect in early B-cell lymphopoiesis. In contrast, CD5 (Ly-1) B cells, a second B-cell population displaying a characteristic set of cell surface markers which are derived from distinct precursors in the peritoneum³⁻⁷, are represented at normal frequencies in these transgenic mice. Thus, the presence of the rearranged immunoglobulin heavy-chain transgene in M54 mice results in an unexpected selective developmental defect that impairs the development of bone-marrow-derived pre-B and B cells without affecting Ly-1 B cells.

M54 transgenic mice produce and secrete substantial amounts of endogenously encoded (Igh-6b allotype) antibodies of all isotypes^{1,2}. They also produce clearly detectable levels of transgene-encoded (Igh-6a allotype) immunoglobulin M (IgM) heavy chains containing a NP^a idiotype variable region which is specific for the haptens 4-hydroxy-3-nitrophenyl (NP)^{1,2,8,9}. Two striking observations have been made concerning immunoglobulin gene expression in these animals^{2,9}. First, a large proportion of the antibodies produced by hybridomas derived from M54 mice express endogenous heavy-chain variable region (V_H) genes that are rarely expressed in normal C57B1/6 mice⁹ and display idiotypes which cross-react with the NP^a idiotype encoded by the transgene. Secondly, rearrangement of endogenous immunoglobulin heavy-chain genes is partially blocked; unrearranged endogenous Igh alleles exist in 40% of pre-B-cell clones derived from M54 bone marrow by transformation with Abelson murine leukaemia virus (A-MuLV)².

These findings led us to inquire whether individual B-cell populations are affected to different extents by the presence of the μ transgene. Two such B lymphocyte subsets have been distinguished by their characteristic cell surface markers and functional properties³⁻⁶. Conventional B cells have intermediate to high levels of surface IgM and IgD, high levels of B220/6B2 (Ly-5) but lack the CD5 (Ly-1) and MAC-1 antigens. Cells of this surface phenotype are repopulated from progenitors in the bone marrow and produce most of the commonly studied anti-

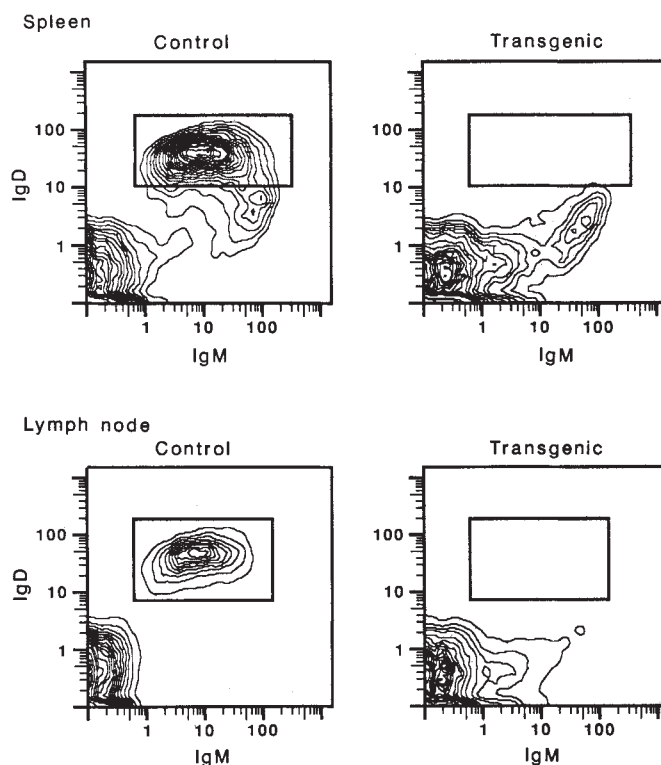


Fig. 1 Conventional B cells are depleted in the spleen and lymph nodes of IgM heavy-chain transgenic (M54) mice. Cells from spleen and peripheral lymph nodes of six-month-old transgenic mice and normal littermates were examined by fluorescence activated cell sorter (FACS) analysis. The box in each panel delineates the conventional (bright IgD) B cell population. Spleen and lymph node cells were prepared as previously described and stained with monoclonal antibodies: anti-IgM, 331.12/FITC; and anti-IgD (Igh-5b), AF6-122/biotin; or anti-B220, RA3-6B2/biotin; or anti-MAC-1; MAC-1/biotin¹⁴. Biotin-conjugated antibodies were revealed with Texas Red/avidin. Purification and fluorochrome conjugation of the monoclonal antibodies has been described¹⁵. The anti-IgM (331.12) recognizes both the Igh-6a and Igh-6b alleles of IgM¹⁶. Dead cells were excluded by propidium iodide staining¹⁷. FACS analyses were conducted as previously described⁷. Data are presented as 5% probability contour maps¹⁸.

body responses in normal mice^{3,4}. Ly-1 B cells, in contrast, have high levels of IgM, low levels of IgD, and express the B220/6B2, MAC-1 and CD5 (Ly-1) surface antigens (A.M.S., unpublished data)³. These cells appear to be self-maintaining in the peripheral immune system in that they are repopulated from immunoglobulin-bearing progenitors present in the peritoneum and thus have been suggested to comprise a separate B-cell lineage⁴⁻⁶. CD5 (Ly-1) B cells produce the majority of auto-antibodies^{3,5,7} and a large proportion of the serum immunoglobulin in normal animals³. Furthermore, they apparently tend to use a restricted set of V_H genes that includes the V_H genes expressed in the hybridomas derived from the transgenic mice^{3,10}.

To determine the proportions of individual B-cell types within lymphatic tissues of M54 transgenic mice, we examined cell-surface marker expression using fluorescent antibodies and flow cytometry with the fluorescence activated cell sorter (FACS). Table 1 and Fig. 1 show that the B cells bearing high levels of surface IgM and IgD (IgM^+IgD^{bright}) are virtually absent from the spleen and lymph nodes of M54 mice. In normal mice, these cells comprise the major fraction of conventional B cells^{3,5}. The few B cells detectable in the transgenic spleen and lymph nodes express intermediate to high levels of surface IgM but have little or no surface IgD (see Fig. 1). The lineage relationship of these

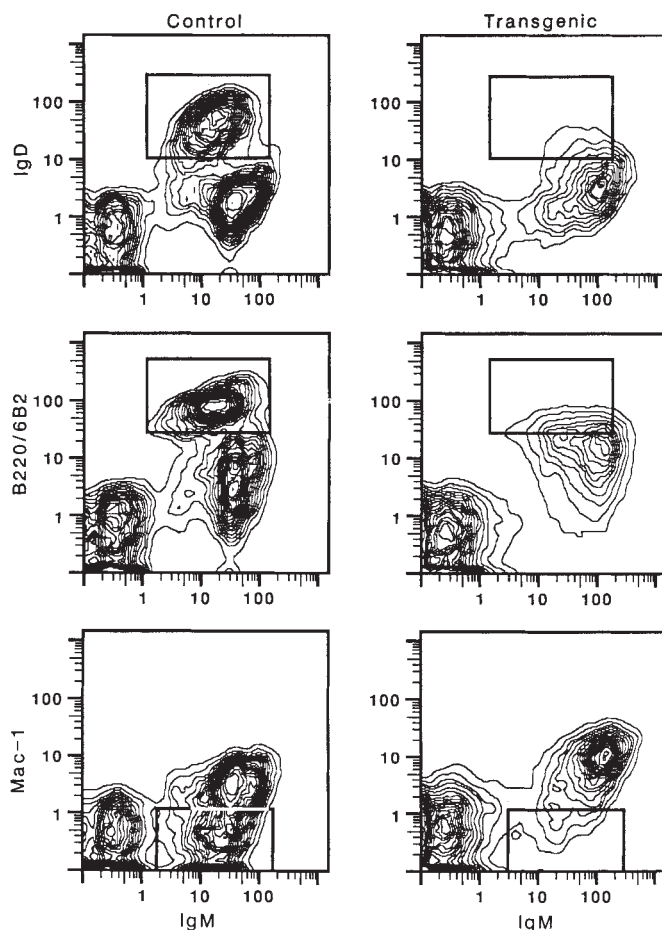


Fig. 2 Conventional B cells are selectively depleted in the peritoneum of IgM heavy-chain transgenic (M54) mice. The cell-surface phenotype of peritoneal B cells from six-month-old transgenic mice and normal littermates was characterized by FACS analysis as described in Fig. 1. The box in each panel delineates the conventional B-cell population.

cells, which express a surface phenotype rarely seen in normal mice, remains to be established.

To quantify CD5 (Ly-1) B-cell representation, we analysed peritoneal cell suspensions from control C57B1/6 and transgenic M54 mice because CD5 (Ly-1) B cells are present at substantially higher frequencies in the peritoneum than in the spleen^{3,4}. Table 1 and Fig. 2 show that cells bearing high levels of surface IgM and little or no surface IgD ($\text{IgM}^+ \text{IgD}^{\text{dull}}$) are represented at normal frequencies in the peritoneum of transgenic mice, whereas cells of the $\text{IgM}^+ \text{IgD}^{\text{bright}}$ surface phenotype, characteristic of conventional B cells, are virtually missing. To demonstrate that the $\text{IgM}^+ \text{IgD}^{\text{dull}}$ cells are not conventional B cells exhibiting an IgD^{dull} surface phenotype due to an allelic exclusion of the endogenous μ/δ heavy-chain locus by the μ transgene, we used two other criteria for the characterization of these cells. First, we found that the levels of expression of the B220/6B2 and MAC-1 cell surface markers are characteristic of CD5 (Ly-1) B cells (Fig. 2). This combination of cell surface antigens has proven more diagnostic for members of the CD5 (Ly-1) B-cell population than the presence or absence of the CD5 (Ly-1) antigen itself because a fraction of CD5 (Ly-1) B cells do not express the CD5 (Ly-1) antigen³. In the transgenic mice studied here, the CD5 (Ly-1) antigen was expressed on the surface of 50% of the peritoneal $\text{IgM}^+ \text{IgD}^{\text{dull}}$ cells (data not shown). Second, FACS studies with anti-IgM allotype reagents that distinguish between transgene-encoded (Igh-6a allotype) and endogenously encoded (Igh-6b allotype) IgM

demonstrated that 80% of peritoneal B cells of M54 mice co-express both types of μ heavy chains (A.S. *et al.*, manuscript in preparation). Thus, the presence of the μ transgene in M54 mice leads to a selective deficiency in the number of conventional B cells without affecting CD5 (Ly-1) B cells. These findings were confirmed by analysis of two additional M54 mice, and similar results were obtained with several mice from the independently derived M95 and M52 mouse lines which carry the same μ transgene (data not shown)¹.

T-cell populations in the transgenic animals were characterized in spleen, lymph nodes and peritoneum by determining the proportions of cells bearing CD4 and CD8 (L3T4 and Ly-2) surface markers. Both T-cell subsets were represented at approximately normal frequencies (Table 1). The cells appeared

Table 1 Numbers of total and individual lymphocyte populations in M54 transgenic and C57B1/6 control mice

	Total lymphocytes recovered	IgM^+ B cells Bright IgD	IgM^+ B cells Dull IgD	T cells CD4 ⁺	T cells Lyt-2 ⁺
Peritoneum					
Control	12	4.0	4.8	0.6	0.6
Transgenic	8	0.2	4.3	0.8	0.7
Spleen					
Control	147	63.2	21.0	27.0	15.0
Transgenic	45	0.5	9.0	12.0	12.0
Lymph node					
Control	8	2.8	0.2	2.3	2.0
Transgenic	16	0.2	1.6	6.1	7.0

Numbers of lymphocytes ($\times 10^6$) were obtained from integrations of FACS analyses and represent the average of two mice. Cells were stained with monoclonal antibodies: anti-IgM, 331.12/FITC and anti-IgD, AF6-122/biotin or anti-CD8 (Lyt-2) 53-7.8/FITC and anti-CD4 (L3T4), GK-1.5/biotin. Staining of cells and FACS analyses were carried out as described in Fig. 1.

phenotypically identical to T cells in control mice by FACS analysis (data not shown). Thus, the lymphopoiesis defect in M54 transgenic mice does not affect the lymphoid T-cell lineage.

We also examined the B cells in the bone marrow of M54 mice, to determine the proportions of the B cell subsets. Conventional bone marrow pre-B cells are detectable either as B220⁺, IgM^- cells¹¹ or as cells that generate relatively small forward and wide-angle light scatter signals (A.M.S., unpublished observation). The number of typical pre-B cells in transgenic bone marrow was reduced sixfold (Fig. 3). Thus, the presence of the transgene seriously diminished the steady-state level of the immediate precursor cell type for most or all normal conventional B cells.

The mechanism responsible for the selective interference with bone marrow lymphopoiesis in these mice is unclear. The depletion of conventional B cells in M54 mice could be due to a rapid turnover of these cells or their progenitors, or to a direct block in B-cell development. The underlying mechanisms for either kind of defect could be intercellular, via interaction with T lymphocytes, or independent of cell-cell interactions. The marked deficiency of pre-B cells in the bone marrow suggests that the transgene exerts its influence through an intracellular mechanism perhaps related to the blockade of endogenous gene rearrangements in conventional B-cell precursors previously described². If so, CD5 (Ly-1) B cells may be less sensitive to the presence of the μ transgene, perhaps because the strategy for gene rearrangement and allelic exclusion used by these cells is fundamentally different from that used by conventional B cells. This hypothetical explanation is consistent with experimental evidence demonstrating V_H segment replacement into rearranged V_{DJ} gene units in differentiating CD5 (Ly-1) B tumour cells¹².

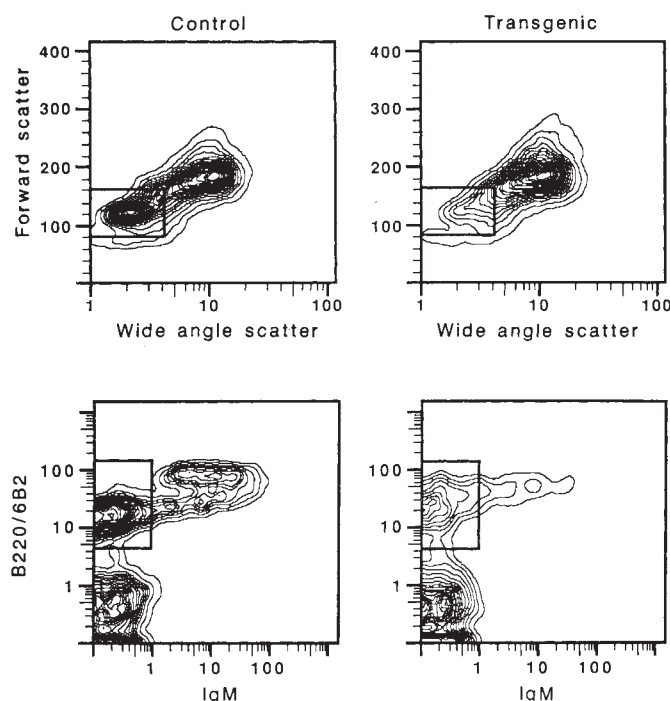


Fig. 3 Pre-B cells are depleted in the bone marrow of IgM heavy-chain transgenic (M54) mice. Bone-marrow cells from six-month-old mice were stained as in Figs 1 and 2. The top panels present the forward angle versus wide angle light scatter of total bone-marrow cells. Pre-B cells have the low forward and wide angle light scatter shown in the boxes in the top panels. The bottom panels present the IgM versus B220/6B2 profiles of the cells within the scatter gates (boxes) shown in the top panels. The boxes in the lower panels delineate the IgM⁻, B220⁺ pre-B cells which constitute 12% of the 6.7×10^7 control and 2.5% of the 5.1×10^7 transgenic bone-marrow cells.

The selective lack of conventional B cells in the transgenic mice potentially explains several anomalies in the previous data. For example, we noted that the transgenic mice gave rise to pre-B lymphoid A-MuLV transformants that all express the transgene and have partially inhibited rearrangement of their endogenous immunoglobulin heavy-chain genes². By contrast, the B-cell hybridomas derived from transgenic mice often do not express the transgene but do express endogenous heavy chain genes⁹. This situation seemed paradoxical because the A-MuLV transformants were thought to represent the precursors of the cells that gave rise to the hybridomas. It now seems likely that in the transgenic M54 mice the A-MuLV transformants represent the conventional B-cell precursors that are unable to mature while the hybridomas mainly represent the CD5 (Ly-1) B-cell population. Other anomalies in the immunoglobulin repertoire of the transgenic mice, particularly the usage of a restricted set of V_H genes for the synthesis of antibodies by transgenic B cell hybridomas⁹, may also be a consequence of the predominance of CD5 (Ly-1) B cells in these animals but interpretations must await more detailed analysis.

We thank P. A. Lalor and D. M. Tarlinton for useful discussions and F. Timothy Gadus for expert technical assistance. This work was supported in part by the NIH. A.M.S. was supported by the USPHS. R.G. is a Scholar of the Leukemia Society of America.

Received 24 February; accepted 14 July 1987.

- Grosschedl, R., Weaver, D., Baltimore, D. & Costantini, F. *Cell* **38**, 647-658 (1984).
- Weaver, D., Costantini, F., Imanishi-Kari, T. & Baltimore, D. *Cell* **42**, 117-127 (1985).
- Herzenberg, L. A. *et al. Immun. Rev.* **93**, 81-102 (1986).
- Hayakawa, K., Hardy, R. R., Parks, D. R. & Herzenberg, L. A. *J. exp. Med.* **161**, 1554-1568 (1985).

- Hayakawa, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 2494-2498 (1984).
- Hayakawa, K., Hardy, R. R., Stall, A. M. & Herzenberg, L. A. *Eur. J. Immun.* **16**, 1313-1316 (1986).
- Hayakawa, K., Hardy, R. R., Parks, D. R. & Herzenberg, L. A. *J. exp. Med.* **157**, 202-218 (1983).
- White-Scharf, M. & Imanishi-Kari, T. *Eur. J. Immun.* **11**, 897-904 (1981).
- Weaver, D. *et al. Cell* **45**, 247-259 (1986).
- Monestier, M. *et al. J. clin. Invest.* (in the press).
- Coffman, R. L. *Immun. Rev.* **69**, 5-23 (1982).
- Kleinfeld, R. *et al. Nature* **311**, 727-733 (1984).
- Hardy, R. R., Hayakawa, K., Parks, D. R. & Herzenberg, L. A. *J. exp. Med.* **159**, 1169-1188 (1985).
- Hardy, R. R. in *Handbook of Experimental Immunology*, 4th edn. (eds Weir, D. M., Herzenberg, L. A., Blackwell, C. C. & Herzenberg, L. A.) 31.1-31.11 (Blackwell, London, 1986).
- Parsons, M., Herzenberg, L. A., Stall, A. M. & Herzenberg, L. A. in *Handbook of Experimental Immunology* 4th edn (eds Weir, D. M., Herzenberg, L. A., Blackwell, C. C. & Herzenberg, L. A.) 97.1-97.17 (Blackwell, London, 1986).
- Loken, M. R. & Stall, A. M. *J. immun. Meth.* **R85-95** (1982).
- Moore, W. A. & Kautz, R. A. in *Handbook of Experimental Immunology* 4th edn (eds Weir, D. M., Herzenberg, L. A., Blackwell, C. C. & Herzenberg, L. A.) 30.1-30.11 (Blackwell, London, 1986).

Phorbol ester and Epstein-Barr virus dependent transformation of normal primary human skin epithelial cells

L. David Tomei*, Ingeborg Noyes*, Denise Blocker†, Jane Holliday & Ronald Glaser*†‡

* Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio 43210, USA

† Department of Medical Microbiology and Immunology, The Ohio State University College of Medicine, Columbus, Ohio 43210, USA

Substantial evidence has implicated Epstein-Barr virus (EBV) in the aetiology of two human neoplasms, nasopharyngeal carcinoma (NPC) and Burkitt's lymphoma^{1,2}. This is supported by the presence of high antibody titres to EBV early antigen and virus capsid antigen^{3,4}, as well as antibody to two viral-associated enzymes, DNase and DNA polymerase^{5,6}. Patients with NPC, particularly the undifferentiated form, are commonly found to have EBV DNA in the tumour². Ito and others⁷⁻¹⁰ have presented strong epidemiological evidence that phorbol esters are related to the unusual geographic distribution of NPC in southeastern regions of China. There appears to be a close link between the widespread EBV infection of the Asian population and the distinct regional distribution in China of plants that produce diterpene ester¹¹. Naturally occurring phorbol esters are produced by plants of the Euphorbiaceae and Thymelaeaceae, which are used as traditional herbal medicines¹². Although it has been established that EBV can infect epithelial cells isolated from NPC as well as certain normal epithelial cells¹³⁻¹⁵, there has been no *in vitro* evidence that EBV induces neoplastic transformation in normal human epithelial cells with or without exposure to phorbol esters. We report here evidence that transformation of normal human epithelial cells results from exposure to infectious EBV and that transformation is dependent on the presence of phorbol esters.

In this study, cell cultures were exposed to 100-fold concentrates of the B95-8 transforming strain of EBV prepared as previously described¹⁶. Cultures received three consecutive virus inoculations (10^8 transforming units per 75 cm² flask containing $\sim 10^6$ cells) at weekly intervals, beginning three weeks after seeding of the epidermal cells. One week after the final virus addition the cells were prepared for soft-agar cloning¹⁷. Primary human skin epidermal cells normally yield 0-50 colonies per 100,000 cells in soft-agar growth medium and quantitation of transformation was based on a substantially enhanced ability to form proliferating colonies in soft-agar growth medium¹⁸.

The active mouse skin tumour promoters 12-O-tetradecanoyl-phorbol-13-acetate (TPA) and phorbol didecanoate (PDD), and

‡ To whom correspondence should be addressed.

Table 1 The effect of EBV in combination with phorbol esters on soft-agar colony formation by primary human epidermal cells

Experiment no.	Solvent alone	No. of soft-agar colonies (per 100,000 cells) EBV alone	Phorbol ester	EBV with phorbol ester
Group A (3 nM TPA used)				
1	9.2 (3.0)	18 (6.0)	—	959 (83.6)
2	0	43 (19.6)	5.5 (2.6)	542 (1.6)
Group B (3 nM PDD used)				
1	30 (3.3)	96 (4.7)	30 (4.1)	404 (45.9)
2	16 (2.8)	246 (57.3)	13 (4.3)	445 (46.0)
Group C (3 nM 4-O-methyl TPA used)				
1	45 (4.4)	76 (12.1)	44 (9.8)	75 (10.4)

Primary cultures of normal human skin epithelial cells were established and maintained using a modification of the method in ref. 29. In this procedure, dispersed keratinocytes were obtained by placing neonatal foreskin specimens into Hank's balanced salt solution containing 0.1% trypsin and 0.02% EGTA followed by blunt dissection of the dermis and epidermis, carried out at the unusually low pH of 3.5 (ref. 17). Cells were seeded into 25-cm² flasks at ~1–3 million cells per flask (yield from each specimen) in MEM medium (pH 7.2) containing 0.7 mM Ca²⁺ and 10% (v/v) heat inactivated fetal bovine serum. The cells were incubated at 37 °C in a humidified atmosphere containing 4% CO₂ and the growth medium was renewed every 2–3 days. Beginning at day 14, 3 nM phorbol ester or an equivalent volume of acetone solvent alone was added to the growth medium and the appropriate cultures were inoculated with B95-8 EBV three times at weekly intervals. Seven days after the last inoculation the cells were trypsinized, dispersed, and prepared for soft-agar (0.3%) colony assay using a modification of a method previously described¹⁸. Replicate 20-cm² plates (*n* = 4) received 100,000 viable cells each and the plates were counted at day 15. Phorbol esters were not added to the soft agar growth medium. Colonies larger than 60 µm in diameter were counted and the final results are expressed in terms of the mean number of colonies per plate with the standard error of the mean in parentheses.

the inactive structural analogue 4-O-methyl-TPA were evaluated. Control cultures received an equivalent volume of solvent alone (5 µl acetone per 5 ml growth medium). The results revealed that primary human epithelial cells that received either TPA or PDD treatment alone or EBV exposure alone did not have significantly increased colony-forming ability in soft agar (Table 1). However, cells that were treated with 3 nM TPA or PDD in combination with EBV exposure showed a marked increase in colony-forming ability, from ~0–50 colonies (normal background) per 100,000 cells, to 500–1,200 colonies per 100,000 cells. Cells treated with 3 nM 4-O-methyl-TPA in combination with EBV did not show enhanced agar colony formation.

Experiments using EBV inactivated by either heat (56 °C for 30 min) or exposure to ultraviolet radiation (1.2 J cm²) showed that cultures treated with inactivated virus did not yield cells having significantly increased agar colony-forming ability as compared with control cultures (Table 2). These data indicate that epithelial cell transformation marked by increased soft-agar

Table 2 The effect of inactivation of EBV on EBV-associated transformation of primary normal human skin epithelial cells

Virus preparation and treatment	Soft-agar colonies per 100,000 cells
Solvent treatment alone, no virus	23 (12)
EBV + 3 nM TPA	504 (116)
Heat-inactivated EBV + TPA	51 (22)
UV-irradiated EBV + TPA	55 (9)

Aliquots of single virus stocks were inactivated using heat (56 °C for 30 min) or irradiation with ultraviolet radiation (1.2 J at 20,000 erg s⁻¹ cm⁻²). Values in parentheses are s.e.m.

Table 3 The effect of different phorbol ester treatment schedules on the expression of EBV-associated transformation

Treatment schedule	TPA (%)	PDD (%)
Phorbol ester added at time of EBV inoculation	100 (6)	100 (11)
Phorbol ester removed at time of EBV inoculation	22 (12)	42 (3)
Continuous exposure to phorbol ester	62 (13)	113 (7)

Phorbol esters were added at the time of cold trypsinization of tissue and renewed every 2–3 days thereafter. Results were compared with those obtained using the standard schedule described in Table 1. Values are percentages, with s.e.m. in parentheses.

colony formation is dependent on infectious virus and an intact viral genome.

The concept of tumour promotion is based on the widely accepted multistage view of carcinogenesis¹⁹. This hypothesis specifies that the promotion step or steps occur after the primary initiation event (in this instance, exposure to EBV). Experiments were performed using three different drug treatment schedules: continuous phorbol ester exposure beginning at the time of trypsin-EGTA treatment and ending when the cells were prepared for agar colony testing; pretreatment alone, where the phorbol ester was removed from the growth medium at the time of EBV addition; and post-treatment alone where the phorbol ester was added to the growth medium after EBV addition. Generally, the least effective treatment schedule was that of pretreatment, whereas post-treatment was equivalent to that of continuous exposure (Table 3). These observations are in accord with the hypothesis that phorbol esters function, at least in this model, as promoters of EBV-associated transformation.

After outgrowth of cell colonies in soft agar, confirmation of the epithelial origin of the colony-forming cells was obtained using the AE1/AE3 monoclonal antibodies against human epithelial keratin (Boehringer Mannheim). All colonies that were isolated and examined were positive for these keratins, indicating the strong likelihood that the cells were of epithelial origin (Fig. 1)²⁰.

Although transformation of the skin epithelial cells was dependent on infectious virus, immunofluorescence analysis did not reveal the presence of EB nuclear antigen (EBNA), early antigen (EA), or virus capsid antigen (VCA). There were not enough cells in the colonies to determine whether EBV DNA was present, because transformation of foreskin epithelial cells following EBV and TPA exposure did not establish an indefinite life-span in culture.

These results indicate that normal primary human epithelial cells can be induced to grow in soft agar after exposure to EBV (a phenotypic change associated with neoplastic transformation) and that the expression of transformation is dependent on the presence of physiological concentrations of phorbol ester tumour promoters. The mechanism of transformation is not clear. However, these observations are consistent with existing knowledge regarding the involvement of EBV and phorbol esters in the aetiology of human neoplasms.

The fact that neither EBV receptors nor virus replication could be demonstrated in these cells is of particular interest. Data from previous studies have shown that only a small number of epithelial cells express EBV receptors or antigens after infection^{13–15,21}. Therefore, if similar frequencies are to be expected, then immunofluorescence analysis of agar colonies in these studies would not permit examination of a sufficiently high number of cells for a reliable determination. Another possibility is that one or more of these EBV antigens may not be expressed at all. Such a phenomenon has been observed in cells microinjected or transfected with EBV DNA (refs 22–24).

There is, however, a third possibility. Failure to detect virus

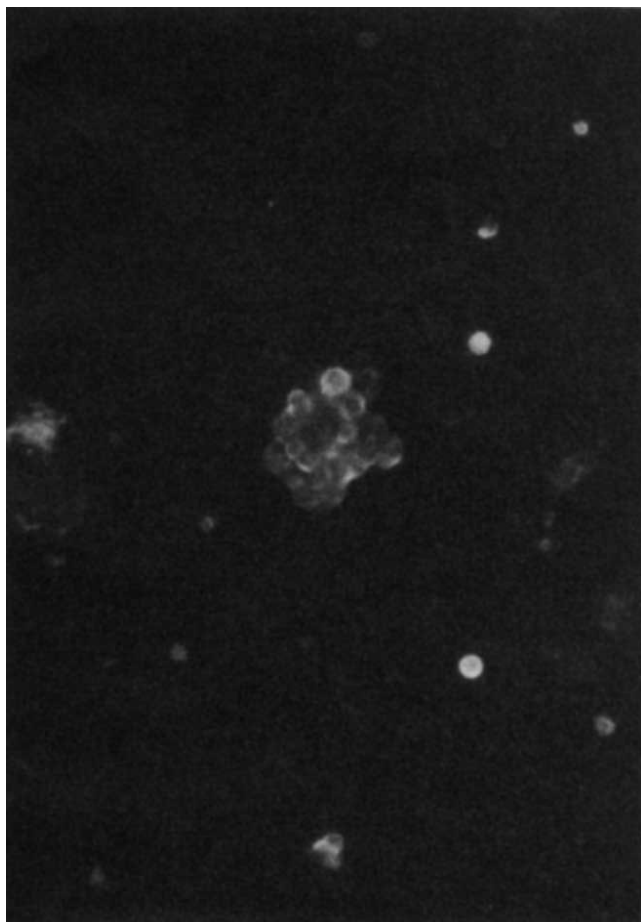


Fig. 1 Detection of keratins in a colony of primary human skin epithelial cells isolated from soft-agar culture following exposure to EBV and TPA as described in Table 1. Cell colonies were removed from soft-agar and fixed with cold acetone/methanol (1:1) for 10 min and air dried on glass slides. They were then stained by indirect immunofluorescence using mouse antihuman monoclonal antibodies AE1 and AE3 and antimouse FITC (fluorescein isothiocyanate)-linked immunoglobulin G (IgG) following the technique recommended by the manufacturer (Hybritech). Magnification $\times 240$ (approx.)

antigens in epithelial cells after EBV-dependent transformation may be related to a virus function or functions yet to be identified. There are three histological types of NPC according to the World Health Organization (WHO) scheme²⁵. These are: keratinizing squamous cell carcinoma (SCC), non-keratinizing carcinoma (NKC), and the undifferentiated form (UC). Although there is lack of agreement among pathologists regarding the existence of three distinct tumour types, there are data that suggest that SCC and UC are separate diseases of the nasopharynx²⁶⁻²⁸. Several laboratories have reported that EBV DNA is associated primarily with UC, and that UC patients have elevated EBV antibody titres^{26,27}. Although the presence of antibodies to EBV EA and VCA has been shown to be of value in diagnosis of UC, it has not been useful for the other histological types²⁶⁻²⁸. Therefore, it is possible that the *in vitro* epithelial cell model described in this study is more closely associated with SCC (not UC) of the nasopharynx in humans, and that there may be a less conventional relationship between SCC and EBV such that the virus could mediate initiation of a neoplastic transformation event without the continued residence of an EBV genome in the tumour cells.

These data support the concept that naturally occurring plant products found in southeastern regions of China may be linked

to the aetiology of NPC in the inhabitants of this region who are also infected with EBV.

We thank Ms A. Beynen and members of the clonogenic assay laboratory for their help. This research was supported in part by a grant from the US Environmental Protection Agency, by a USPHS grant, and by the PHS grant to the Comprehensive Cancer Center awarded by the NCI, DHHS.

Received 27 April; accepted 14 July 1987.

1. Epstein, M. A., Achong, B. G. & Barr, Y. M. *Lancet* **i**, 702-703 (1964).
2. Zur Hausen, H. *et al. Nature* **228**, 1056-1058 (1970).
3. Henle, W. *et al. J. natn. Cancer Inst.* **44**, 225-231 (1970).
4. Henle, G., Henle, W. & Klein, G. *Int. J. Cancer* **8**, 272-282 (1971).
5. Cheng, Y. C., Chen, J. Y., Glaser, R. & Henle, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6162-6165 (1980).
6. Chiou, J. F., Li, J. K. K. & Cheng, Y. C. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5728-5731 (1985).
7. Hirayama, T. & Ito, Y. *Prev. Med.* **10**, 614-622 (1981).
8. Ito, Y., Kawanishi, M., Harayama, T. & Takabayashi, S. *Cancer Lett.* **12**, 175-180 (1981).
9. Mizuno, F. *et al. Cancer Lett.* **19**, 199-205 (1983).
10. Zeng, Y. *et al. Cancer Lett.* **23**, 53-59 (1984).
11. Fedder, M. & Gonzalez, M. F. *Am. J. Med.* **79**, 365-369 (1985).
12. Duke, J. A. & Ayensu, E. S. in *Medicinal Plants of China* 300-314 (Reference Publications, Algonac, Michigan, 1985).
13. Glaser, R., de-The, G., Lenoir, G. & Ho, J. H. C. *Proc. natn. Acad. Sci. U.S.A.* **73**, 960-963 (1976).
14. Glaser, R. *et al. J. natn. Cancer Inst.* **64**, 1085-1090 (1980).
15. Sixbey, J. W., Nedrud, J. G., Raab-Traub, N., Hanes, R. A. & Pagano, J. S. *New Engl. J. Med.* **310**, 1225-1230 (1984).
16. Stoerker, J., Yajima, Y. & Glaser, R. *Int. J. Cancer* **30**, 385-392 (1982).
17. Lehman, T. A., Noyes, I. & Milo, G. E. *Tissue Culture Association Manual* (in the press).
18. Hamburger, A. & Salmon, S. E. *Science* **197**, 461-463 (1977).
19. Farber, E. *Cancer Res.* **44**, 4217-4223 (1984).
20. Shi, S.-R. *et al. Am. J. Path.* **117**, 53-63 (1984).
21. Young, L. S., Clark, D., Sixbey, J. W. & Rickinson, A. B. *Lancet* **i**, 240-242 (1986).
22. Graessman, A., Wolf, H. & Bornkamm, G. W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 433-436 (1980).
23. Stoerker, J., Parris, D., Yajima, Y. & Glaser, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5852-5855 (1981).
24. Stevenson, M., Volsky, B., Hedenskog, M. & Volsky, D. J. *Science* **233**, 980-984 (1986).
25. Shanmugaratnam, K. & Sabin, L. H. *International Histological Classification of Tumors No. 19* (World Health Organization, Geneva, 1978).
26. Andersson-Anvret, M., Forsby, N., Klein, G. & Henle, W. *Int. J. Cancer* **20**, 486-491 (1977).
27. Neel, H. B. III *et al. Laryngoscope* **90**, 1981-1990 (1980).
28. Michéau, C., de-The, G., Schwaab, G. & Lengrand, D. in *Nasopharyngeal Carcinoma: Current Concepts* (eds Prasad, U. *et al.*) 65-70 (University of Malaya, Kuala Lumpur, 1983).
29. Kitano, Y. & Hata, S. *Arch. dermat. Forsch.* **245**, 203-210 (1972).

An intronless gene encoding a potential member of the family of receptors coupled to guanine nucleotide regulatory proteins

Brian K. Kobilka, Thomas Frielle, Sheila Collins, Theresa Yang-Feng*, Tong Sun Kobilka, Uta Francke*, Robert J. Lefkowitz & Marc G. Caron

Howard Hughes Medical Institute, Departments of Medicine, Biochemistry and Physiology, Duke University Medical Center, Durham, North Carolina 27710, USA

* Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA

Plasma membrane receptors for hormones, drugs, neurotransmitters and sensory stimuli are coupled to guanine nucleotide regulatory proteins. Recent cloning of the genes and/or cDNAs for several of these receptors including the visual pigment rhodopsin¹, the adenylate-cyclase stimulatory β -adrenergic receptor²⁻⁴ and two subtypes of muscarinic cholinergic receptors^{5,6} has suggested that these are homologous proteins with several conserved structural and functional features. Whereas the rhodopsin gene consists of five exons interrupted by four introns¹, surprisingly the human and hamster β -adrenergic receptor genes contain no introns in either their coding or untranslated sequences⁷. We have cloned and sequenced a DNA fragment in the human genome which cross-hybridizes with a full-length β_2 -adrenergic receptor probe

at reduced stringency. Like the β_2 -adrenergic receptor this gene appears to be intronless, containing an uninterrupted long open reading frame which encodes a putative protein with all the expected structural features of a G-protein-coupled receptor.

We have previously reported⁷ that high-stringency Southern blot analysis of human genomic DNA, performed with a full-length β_2 -adrenergic receptor cDNA probe, reveals only the β_2 -adrenergic receptor gene, present in a single copy. When the analysis is performed at lower stringency an additional weak signal is detected, corresponding to a 3.8-kilobase (kb) *Eco*RI fragment. This fragment was cloned from a size-selected partial genomic library screened at reduced stringency with the full-length human β -adrenergic receptor cDNA clone pTF³. The restriction map of clone G-21 is shown in Fig. 1a and the sequence of the *Xba*I-*Bam*HI fragment is shown in Fig. 1b. When this fragment was labelled and used to probe human genomic DNA digested with different restriction enzymes, the pattern of intense hybridization observed in this blot is identical to that of the weak signals seen on a low-stringency human genomic blot when the β_2 -adrenergic receptor sequence is used as a probe (data not shown). This verifies that G-21 represents a genomic clone for the 3.8-kb *Eco*RI fragment initially observed on low-stringency genomic blots⁷.

The DNA sequence (Fig. 1b) reveals that the *Xba*I-*Bam*HI fragment contains a 1,309-base-pair (bp) open reading frame (ORF). A putative protein beginning with the first methionine in this ORF and extending to the termination codon is 421 amino acids in length. This compares with 418 amino acids for the hamster β_2 -adrenergic receptor and 413 amino acids for the human β_2 -adrenergic receptor. Hydrophobicity analysis of this putative protein reveals seven clusters of hydrophobic residues of >20 amino acids long, comparable to the analyses of the human and hamster β_2 -adrenergic receptors, porcine cerebral and cardiac muscarinic receptors and rhodopsin¹⁻⁶.

A comparison of the amino-acid sequence of the putative G-21 protein with the human β_2 -adrenergic receptor, avian β -adrenergic receptor, porcine cerebral and cardiac muscarinic receptors and human rhodopsin is shown in Fig. 2. Note that the greatest homology between G-21 and these receptors occurs in the putative transmembrane domains. Within these domains G-21 shares 43% homology with the human β_2 -adrenergic receptor, 48% with the avian β -adrenergic receptor, 39% with the human α_2 -adrenergic receptor (Kobilka *et al.*, manuscript in preparation), 31% with the porcine cerebral muscarinic receptor, 29% with the porcine cardiac muscarinic receptor and 22% with human rhodopsin. The single greatest stretch of amino-acid and nucleotide homology occurs between G-21 and the human β_2 -adrenergic receptor in the sixth transmembrane domain. Here there are 19 of 20 amino acids in common and 54 of 60 bp identities. Several amino-acid residues in G-21 are highly conserved in all of the G-protein-coupled receptors cloned thus far. These include two Asp residues in the first and second helices at positions 82 and 116, and prolines in the fourth to seventh hydrophobic domains at positions 168, 169, 206, 359, 395. Several other features of the structure of this putative receptor indicate that it is a member of the family of G-protein-coupled receptors. Cytoplasmic loops I and II (which connect transmembrane domains I and II and III and IV) are particularly well conserved between G-21, β -adrenergic and muscarinic receptors. In contrast to the β_2 -adrenergic receptor, the putative G-21 protein contains a relatively large hydrophilic domain (127 amino acids) between the fifth and sixth hydrophobic putative transmembrane domains and a relatively short C terminus (18 amino acids) after the seventh hydrophobic domain. This pattern is more similar to the muscarinic receptors^{3,6} which are coupled to G proteins regulating phospholipase C, a K⁺ channel and inhibition of adenylate cyclase. Whereas the conserved cytoplasmic sequences of these proteins may be implicated in the common features of their interactions with G proteins, the highly variable regions (for example, the third cytoplasmic loop and

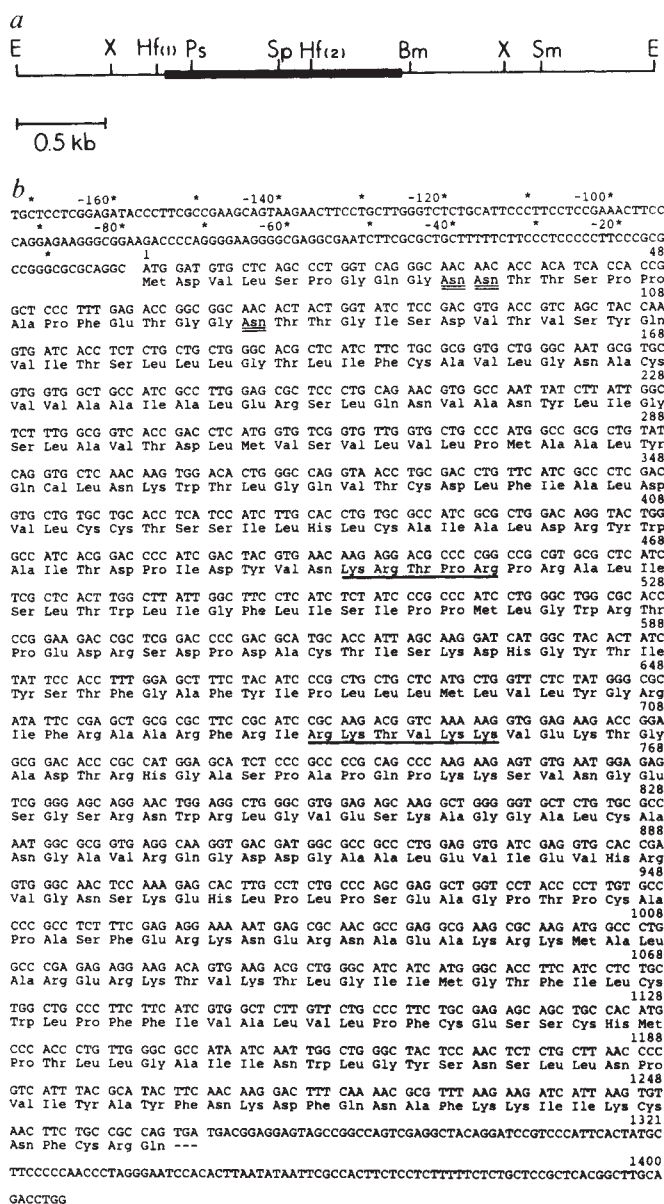


Fig. 1 **a**, Restriction map of the 3.8-kb *Eco*RI clone G-21. The coding region is indicated by the heavy line. Enzymes: Bm, *Bam*HI; E, *Eco*RI; Hd, *Hind*III; Hf, *Hinf*I; Ps, *Pst*I; Sm, *Sma*I; Sp, *Sph*I. **b**, DNA sequence of the coding region of clone G-21. The nucleotide positions are numbered relative to the initiator ATG codon for the putative G-21 protein. The amino-acid residues underlined with a single heavy line are consensus sites for phosphorylation by protein kinase C. The asparagine residues indicated with a double line are potential sites for N-linked glycosylation.

Methods. The 3.8-kb G-21 genomic clone was obtained from a size selected, human partial genomic library. Human genomic DNA was digested with *Eco*RI and fragments of 3–4.5 kb were isolated and cloned into *Eco*RI-cut, phosphatased *ygt*10 arms (Stratagene). The phage DNA was packaged (Gigapack, Stratagene), and the library was screened using human β_2 -adrenergic receptor cDNA clone pTF3 (ref. 8) labelled with ³²P by the random hexamer priming method¹⁷ as a probe. Hybridization was performed at 45 °C in 6×SSC and filters were washed in 2×SSC at 60 °C. Both strands of the coding region were sequenced using the dideoxy chain termination method¹⁸.

C terminus) may dictate the specificity of interaction with the different G proteins.

Chromosomal localization of the G-21 clone was determined by Southern blot analysis of DNA from 12 hamster and human somatic cell hybrids by methods previously described^{8,9}.

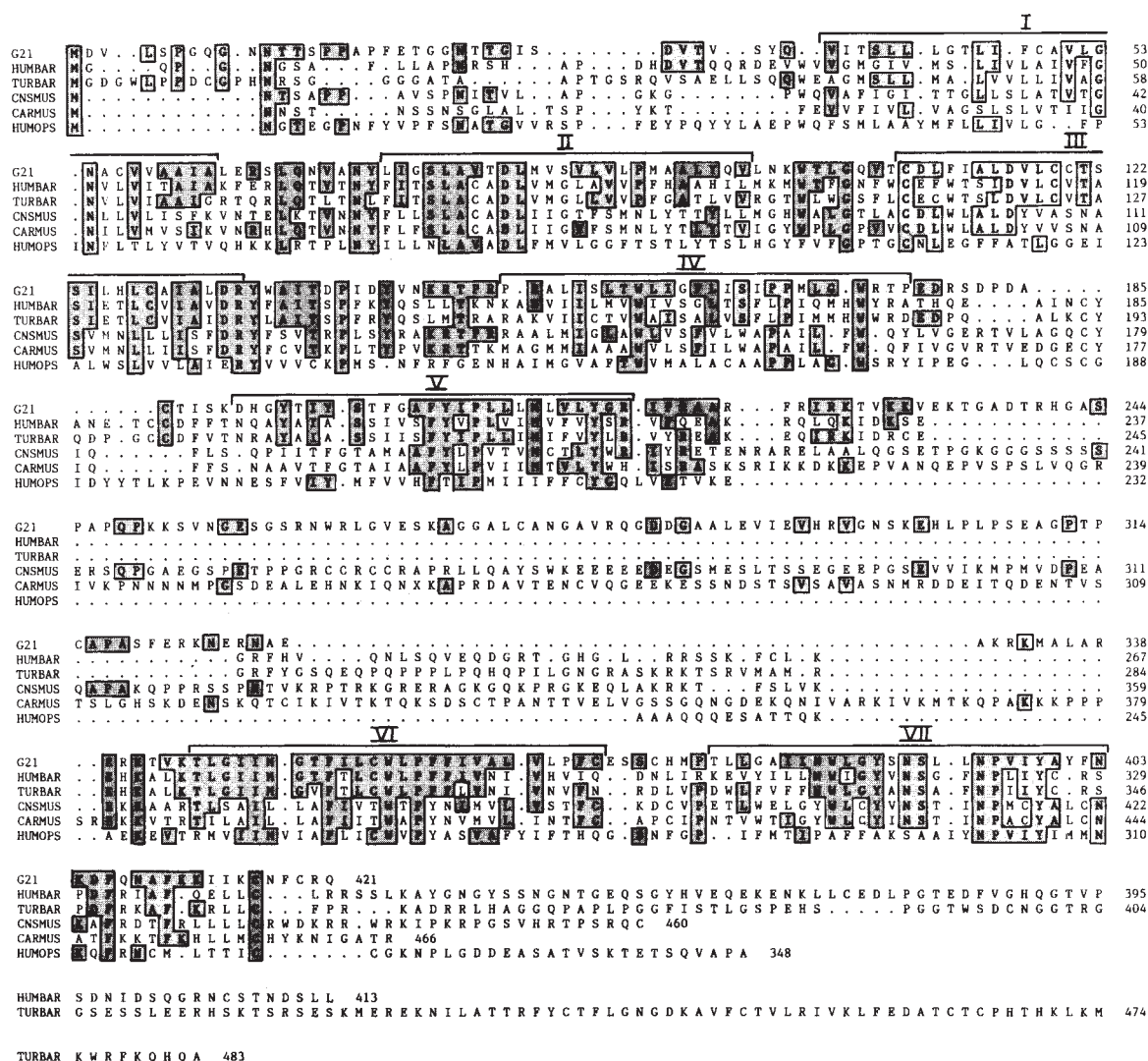


Fig. 2 Alignment of the amino-acid sequences for the putative G-21 protein, the human β_2 -adrenergic receptor (HUMBAR)³ the avian β -adrenergic receptor (TURBAR)⁴, the porcine cerebral muscarinic receptor (CNSMUS)⁵, the porcine cardiac muscarinic receptor (CARMUS)⁶ and human opsin (HUMOPS)¹. Gaps (.) have been inserted to maximize homology. Amino acids are numbered at the end of each line. The numbered brackets indicate potential transmembrane segments based on hydrophobicity analysis. The shaded residues indicate amino-acid identities between G-21 and the other proteins.

Chromosome 5 was the only human chromosome that segregated perfectly with the G-21 gene; every other human chromosome could be excluded by two or more discordant hybrids. *In situ* hybridization of [³H]-labelled G-21 probe to human metaphase chromosomes by previously described methods^{10,11} resulted in specific labelling over bands q11.2-q13 of chromosome number 5. No other chromosomal site was specifically labelled above background. Thus, like the β_2 -adrenergic receptor, G-21 is located on chromosome 5, but at a more proximal locus. The chromosomal position of G-21 is the same as that for the glucocorticoid receptor¹².

Northern blot analysis of 15- μ g samples of total human RNA from fetal tissues and placenta was performed at high stringency (65 °C, 0.2 $\times$ SSC) using a 819-bp probe comprising nucleotides -58-761 of the 5'-untranslated and coding block of G-21. As shown in Fig. 3 a single messenger RNA species of ~6,000 nucleotides was most abundant in lymphatic tissues including lymph node, spleen and thymus. This RNA was also expressed in gut, muscle and kidney but not in lung, heart, adrenal and placenta. By contrast we have previously demonstrated that the size of the β_2 -adrenergic receptor mRNA is ~2,200 nucleotides.

Capped SP6 RNA produced from the G-21 coding sequence can efficiently direct the synthesis of an unglycosylated protein of relative molecular mass (M_r) 42,000 (42K) in reticulocyte lysates in the absence and a 44K membrane-associated protein in the presence of microsomal membranes (Fig. 4). The G-21 protein translated in the presence of microsomal membranes sediments with the membrane fraction (lane 6). These properties are identical to those observed with the β_2 -adrenergic receptor (Fig. 4). Thus the ORF in G-21 is capable of encoding a protein which, like the β_2 -adrenergic receptor, undergoes core glycosylation in the endoplasmic reticulum, but which has no cleaved signal sequence.

SP6 RNA from G-21 was injected into *Xenopus laevis* oocytes, membranes were harvested after 2 days and attempts were made to detect specific binding for various catecholamine receptors using ligands for the β_1 -, α_1 - and α_2 -adrenergic as well as D₁ and D₂ dopamine receptors without success. The specific function of the putative G-21 receptor is thus presently unknown.

Our major interest in cloning and studying this genomic fragment has been to further our understanding of the protein structure as well as the structure and evolution of genes for G-protein-coupled receptors. Like the genes for the mammalian

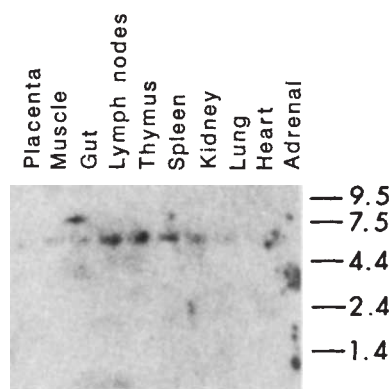


Fig. 3 Northern blot analysis of G-21 transcripts in fetal human tissues and human placenta.

Methods. Total cellular RNA was prepared from each tissue as previously described¹⁹. RNA (15 µg) was denatured in glyoxal²⁰, electrophoresed through 1.2% agarose gels and transferred for Northern blotting as previously described²¹. The filter was hybridized at 42 °C for 24 h in a cocktail containing 50% formamide, 5× SSC, 1× Denhardt's 0.1% SDS, 10% dextran sulphate, 100 µg ml⁻¹ salmon sperm DNA. The probe was derived from the 819-bp *Hinf*I fragment of the G-21 insert by labelling with α[³²P]dCTP by nick translation²². For hybridizations, 4×10⁶ dpm ml⁻¹ was used. Final washing conditions were 0.2× SSC, 0.1% SDS, 65 °C. The filter was exposed to Kodak XAR-5 film for 4 days at -80 °C with two intensifying screens.

β₂-adrenergic receptors, and in contrast to those for the family of visual pigments in mammals, the entire 421-amino-acid coding block for the putative receptor protein is contained within one continuous exon. Whether the gene for G-21 contains introns in its 5' or 3' untranslated regions remains to be established.

It is unlikely that G-21 represents a pseudogene for the β₂-adrenergic receptor or some other gene, for several reasons. Most pseudogenes do not contain uninterrupted coding blocks because of the lack of selective pressure in preventing termination mutations¹³. For the same reason one might expect all regions of a pseudogene to mutate with the same frequency. Thus, one would not expect to find well-conserved regions of homology as we observe between G-21 and the G-protein-coupled receptors. Finally, the G-21 gene is expressed in several tissues as revealed by Northern blot analysis. That the mRNA hybridized is in fact that corresponding to G-21 and not some other cross-hybridization species is indicated by the high stringency of the hybridization. Moreover, the size of the mRNA is considerably larger than that of the β₂-adrenergic receptor but very similar to that for the M₂ muscarinic receptor, to which it is only distantly related structurally. The tissue distribution of the mRNA is unique, being highest in lymphoid tissues. This might indicate that G-21 encodes a receptor for a lymphokine or other substance with specific actions on this type of cell.

If, as appears likely, G-21 is a gene for a member of the class of G-protein-coupled receptors, it raises several interesting and provocative questions regarding the origin and functional significance of introns. Did G-21 and the β₂-adrenergic receptor evolve from the opsin family before the opsins had introns or did they evolve as processed genes for the opsins or some other functionally related intron-containing gene? It has been suggested that introns may in some way enhance the efficiency of mRNA processing in the nucleus^{14,15}. If so how does the lack of introns affect the expression of these genes? It has been speculated that introns are important in increasing genetic diversity¹⁶. Exons often encode functional units of proteins such as membrane-spanning domains, ligand-binding sites or catalytic units. G-protein-coupled receptors also have several functional domains such as those involved in ligand binding, G-protein

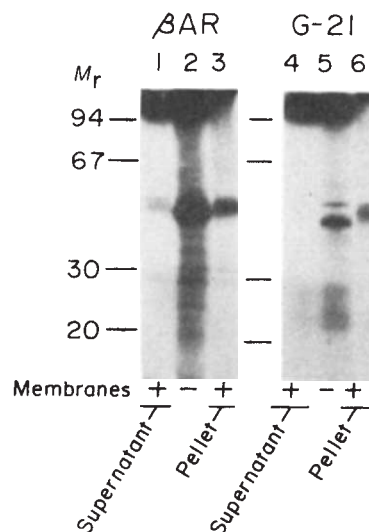


Fig. 4 SDS-polyacrylamide gel electrophoresis and autoradiography of ³⁵S-methionine-labelled protein prepared by *in vitro* translation of capped RNA made from either β₂-adrenergic receptor cDNA or G-21 cloned into an SP6 expression vector. Translations were done with and without canine pancreatic microsomal membranes as indicated at the bottom of each lane. Translation mixtures were centrifuged to separate membrane-associated proteins from soluble proteins. Molecular weight standards are indicated at the left. Lanes 1–3 are translations from the human β₂-adrenergic receptor (pSPTF) (Kobilka *et al.*, manuscript in preparation). In lanes 1 and 3 microsomal membranes were present; the mixture was centrifuged to produce a pellet (lane 3) and a supernatant fraction (lane 1). Lane 2 shows the results when membranes were not present. Lanes 4–6 are the corresponding translations from G-21 (pSPHB3').

Methods. A SP6 transcription template pSPTF was made by cloning the human β₂-adrenergic receptor cDNA (pTF3) into the *Eco*RI site of pSP65 (Amersham) (Kobilka *et al.*, manuscript in preparation). An SP6 expression vector pSPXB was prepared by ligating the *Xba*I–*Bam*HI fragment of G-21 (see Fig. 1a) into these restriction sites in the polylinker of pSP64 (SP6 promoter–*Hind*III–*Pst*I–*Sall*–*Xba*I–*Bam*HI–*Sma*I–*Sac*I–*Eco*RI). The 5'-untranslated region was removed to improve the efficiency of translation by digesting pSPXB with *Hind*III, filling in the *Hind*III end with the Klenow fragment of DNA polymerase, digesting with *Pst*I and isolating the large fragment containing the *Pst*I–*Bam*HI fragment of G-21 attached at the *Bam*HI site to pSP64. This fragment was ligated with the *Hinf*I(1)–*Bam*HI fragment prepared by digesting G-21 with *Hinf*I, filling in the *Hinf*I end with the Klenow fragment of DNA polymerase then digesting with *Pst*I and isolating the fragment from an agarose gel. The resulting construct pSPHB contained the *Hinf*I(1)–*Bam*HI fragment of G-21. To enhance the stability of RNA produced from this vector the 3'-untranslated region of the human β₂-adrenergic receptor containing a poly(A) tail was isolated by digesting pTF3 (ref. 3) with *Eco*RV and *Eco*RI and ligating this fragment into the *Sma*I–*Eco*RI site in pSPHB. The resulting construct pSPHB3' was used in these experiments. RNA was prepared from pSPHB3' and pSPTF by linearizing the DNA with *Eco*RI and *Sall*I respectively and transcribing with SP6 RNA polymerase according to methods previously described²³.

coupling and receptor regulation. But in contrast to most functionally complex molecules, all of these domains are encoded by a single exon in the β₂-adrenergic receptor and in the putative receptor protein represented by G-21. The cloning and characterization of genes for other members of this family of receptors should lead to greater understanding of the evolution of such multifunctional proteins in the absence of introns.

We thank Donna Addison and Mary Holben for assistance in the preparation of this manuscript and Drs Cynthia Morton and Frederick Bieber for providing the fetal human RNA samples. This work was supported in part by the NIH.

Received 15 June; accepted 10 July 1987.

1. Nathans, J. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4851-4855 (1984).
2. Dixon, R. A. F. *et al. Nature* **321**, 75-79 (1986).
3. Kobilka, B. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 46-50 (1987).
4. Yarden, Y. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 6795-6799 (1986).
5. Kubo, T. *et al. Nature* **323**, 411-416 (1986).
6. Kubo, T. *et al. FEBS Lett.* **209**, 367-372 (1986).
7. Kobilka, B. K. *et al. J. biol. Chem.* **262**, 7321-7327 (1987).
8. Baas, F., Bilker, H., Van Ommen, G.-J. B. & de Vijlder, J. J. M. *Hum. Genet.* **67**, 301-305 (1984).
9. de Martinville, B., Wyman, A. R., White, R. & Francke, U. *Am. J. hum. Genet.* **34**, 216-26 (1982).
10. Harper, M. E. & Saunders, G. F. *Chromosoma* **83**, 431-439 (1981).
11. Yang-Feng, T. L., Floyd-Smith, G., Nemer, M., Drouin, J. & Francke, U. *Am. J. hum. Genet.* **37**, 1117-1128 (1985).

12. de la Chapelle, A. (ed.) *Cytogenet. Cell Genet.* **40**, (8th Int. Workshop Gene mapping) (Karger, Basle, 1985).
13. Vanin, E. F. A. *Rev. Genet.* **19**, 253-272 (1985).
14. Hamer, D. H. & Leder, P. *Cell* **18**, 1299-1302 (1979).
15. Volckaert, G., Feunteun, J., Crawford, L. V., Berg, P. & Fiers, W. J. *J. Virol.* **30**, 674-682 (1979).
16. Gilbert, W., Marchionni, M. & McKnight, G. *Cell* **46**, 151-154 (1986).
17. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).
18. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
19. Chirgwin, J. M., Przybyla, A. E., McDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
20. McMaster, G. K. & Carmichael, G. C. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4835-4838 (1977).
21. Aziz, N. & Munro, H. N. *Nucleic Acids Res.* **14**, 915-927 (1986).
22. Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184-1188 (1975).
23. Melton, D. W., McEwan, C., McKie, A. B. & Reid, A. M. *Cell* **44**, 319-328 (1986).

Association of DNA-bound progesterone receptors

Bernard Théveny*, Alain Bailly†, Claudine Rauch†, Michel Rauch†, Etienne Delain* & Edwin Milgrom†‡

* Laboratoire de Microscopie Cellulaire et Moléculaire (U.A. 147 CNRS), Institut Gustave-Roussy, 94805 Villejuif Cedex, France

† Unité 135 INSERM, Faculté de Médecine Paris-Sud, 94275 Le Kremlin-Bicêtre Cedex, France

‡ To whom correspondence should be addressed.

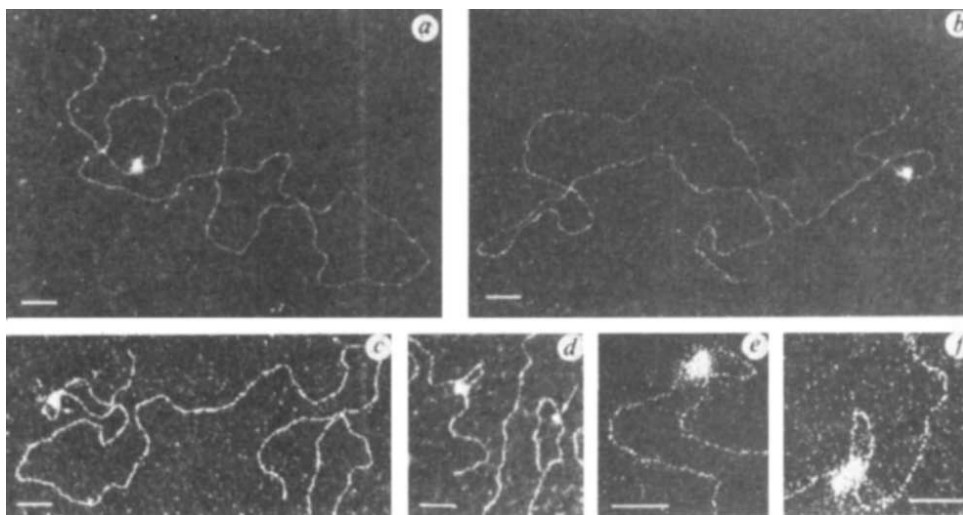
Steroid hormone-receptor complexes regulate the transcription of specific genes. Recent studies of high-affinity interactions between the receptors and discrete regions of DNA, together with gene-transfer experiments, have led to the precise mapping of hormone regulatory elements^{1,2}. Nothing is known, however, about the mechanisms whereby DNA-bound receptors modulate gene transcription. At the start of transcription in prokaryotes two oligomeric molecules of several regulatory proteins must bind to two specific DNA sites and interact with one another to regulate the binding of RNA polymerase to DNA^{3,4}. Using electron microscopy to observe progesterone receptor binding to regulatory regions of uteroglobin and mouse mammary tumour virus genes, we demonstrate a similar binding between receptor oligomers at

two DNA sites. DNA loops are formed when the hormone regulatory elements are at a distance from one another. Thus, in common with certain prokaryotic systems, protein-protein interactions may be important in steroid hormone regulation of gene transcription.

The progesterone receptor binds with high affinity to two discrete regions of DNA, located upstream from the uteroglobin gene. The first site, at position -2427 to -2376, exhibits two DNase footprints and the second, further upstream between nucleotides -2709 and -2620, gives three DNase footprints. Binding sites with lower affinity are found in the first intron of the uteroglobin gene⁵. Immunoaffinity purified rabbit progesterone receptor⁶ has been incubated with a DNA fragment encompassing both regions (nucleotides -3257 to -395) and the receptor-DNA complexes isolated by Superose-6B chromatography. Two types of complex were evident in the electron microscope. In many cases a single receptor oligomer was attached to the DNA (Fig. 1a, b). Measurements showed that the binding occurred at exactly the sites mapped by DNase footprints (not shown). It was not feasible to size the receptor oligomer precisely and thus determine the number of constituent subunits. A second type of complex was observed as a loop consisting of a receptor polymer bound to both sites (proximal and distal) (Fig. 1, c-f). The frequency of these loops was relatively high: we observed 154 DNA molecules having a bound receptor, of which 96 had a receptor oligomer attached to a single site (I or II) and 58 had a receptor at both sites, with 33 (56.9%) forming DNA loops.

Loop formation could arise in two ways: either a single

Fig. 1 Binding of progesterone receptor to DNA upstream from the uteroglobin gene. Receptor oligomers were bound to a, site I (-2427 to -2376) or to b, site II (-2709 to -2620). c, Two receptor oligomers attached to both binding sites to make a DNA loop. d, Two DNA-bound receptor oligomers inducing a DNA loop are shown beside a single oligomer bound to the site I (arrow). e, f, DNA loops at higher magnification. Scale bars equal 50 nm.



Methods. An upstream fragment of the uteroglobin gene (-3257 to -395) was cloned in pBR322. The plasmid DNA was linearized with *EcoRI* and incubated ($10 \mu\text{g ml}^{-1}$) with the purified progesterone receptor ($20 \mu\text{g ml}^{-1}$) for 90 min at 0°C in $40 \mu\text{l}$ reaction buffer (20 mM Tris HCl, pH 7.4, 150 mM NaCl, 2 mM dithiothreitol, 0.1 mM EDTA, 1.8 mg ml^{-1} bovine serum albumin, 0.1 μM unlabelled ORG 2058 [16 α -ethyl-21-hydroxy-19-norpregn-4-ene-3,20-dione] (Amersham)). No fixative agents were added and the complexes were purified by chromatography on a Superose-6B column equilibrated with reaction buffer without BSA. Five μl aliquots of peak fractions (DNA concentration $0.5 \mu\text{g ml}^{-1}$) were deposited on carbon-coated 600 mesh grids activated by glow discharge in the presence of pentylamine¹⁹. Specimens were contrasted with 1% aqueous uranyl acetate, dried and observed without shadowing by annular darkfield electron microscopy²⁰ combined with the inelastic spectroscopic imaging mode of the Zeiss 902 electron microscope²¹. To ensure that the protein bound to DNA was not a contaminant of the receptor preparation two aliquots were chromatographed on immunoaffinity columns containing antireceptor MI 60-10 antibody or a non-receptor related antibody⁶. 68% decrease of protein-DNA complexes was seen only when the MI 60-10 antibody was used.

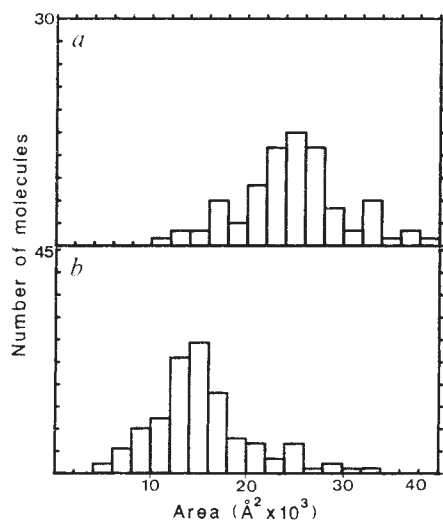


Fig. 2 Size of receptor oligomers bound to a single DNA site, or inducing DNA loops. Measurements of the area corresponding to receptor oligomers for *a*, loop-forming oligomer-DNA complexes: area distribution of 80 loop-forming receptor oligomer-DNA complexes, or *b*, single receptor oligomer-DNA complexes, area distribution of 120 measurements. The highest class position in *a* is approximately twice that for *b*. Measurements were made using an IPS-Kontron Image Analyzer on TV images magnified $\times 650,000$.

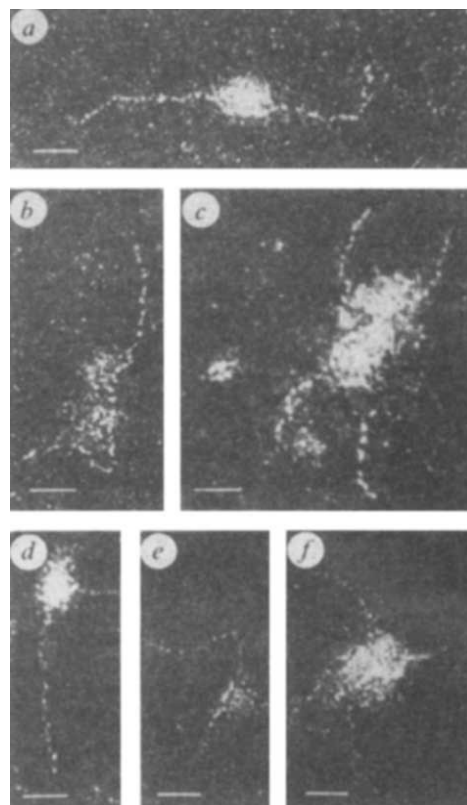


Fig. 3 *a-c*, Oligomer of the progesterone receptor and a 430-base-pair (bp) fragment from the upstream region (-330 to $+100$) of the MMTV gene. This region contains the two steroid receptor binding sites (the receptor-DNA complexes were obtained by the protocol described in Fig. 1). *a*, Represents progesterone receptor oligomer binding to a single site. *b*, Two oligomers binding at two sites present on the same fragment, *c*, the association of two oligomers bridges two different DNA molecules. *d-f*, Interaction between progesterone receptor oligomer and a 302-bp fragment from the upstream region of the uteroglobin gene (-2871 to -2569). This fragment contains a single receptor binding site. *d*, *e*, Receptors bind to the DNA fragment at site II. *f*, Two DNA molecules are linked by progesterone receptor oligomers. Scale bars equal 20 nm.

receptor oligomer could bind DNA at two sites or two DNA-bound receptor oligomers could interact to produce the DNA loop. To distinguish between these possibilities we compared the size of the proteins bridging the loops to the size of the proteins bound to a single DNA site. The average size of the proteins present in the loops was approximately double that of proteins bound to DNA at a single site (Fig. 2), suggesting that the loops were generated by association of two DNA-bound receptor oligomers. In addition, no loops were observed when we used DNA from a mutant plasmid in which the proximal site had been deleted (data not shown).

The mouse mammary tumour virus (MMTV) promoter interaction with glucocorticoid⁷⁻⁹ and progesterone⁹ receptors has been analysed, and the functional significance demonstrated by transfection experiments. The hormone regulatory elements (HREs) of the uteroglobin gene have only recently been studied by transfection (J. F. Savouret, unpublished observations). Therefore, we decided to examine further the binding of the progesterone receptor and a DNA fragment which included the proximal and distal sites of the MMTV gene (-330 to $+100$ nucleotides). Progesterone receptor oligomers were seen to bind

to one of the sites (Fig. 3*a*), or to both adjacent sites (Fig. 3*b*). It was not obvious whether the proteins were associated or simply lying side-by-side on the DNA, but no receptors were seen in a *trans* position. Also, in the presence of relatively high concentrations of DNA, receptors bridging two different DNA

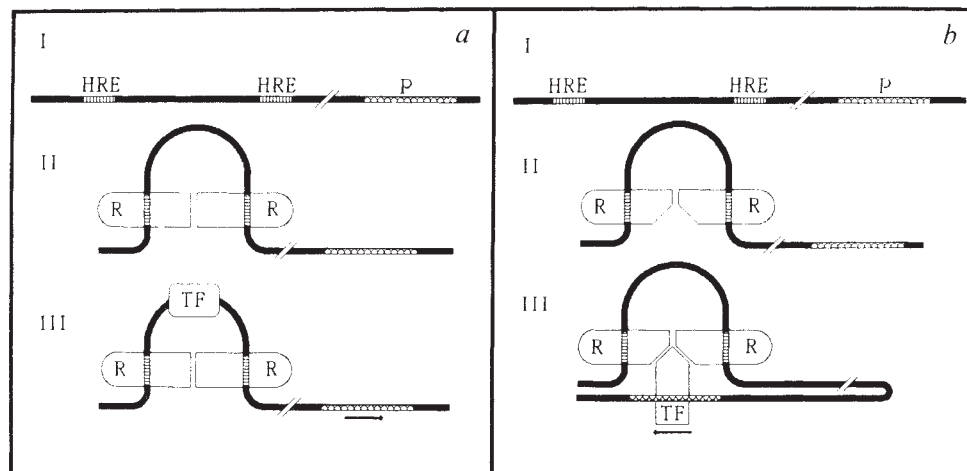


Fig. 4 Possible function of the interaction between DNA-bound receptors in the regulation of gene transcription. *a*, Model in which receptor oligomers (R) bind to both hormone regulatory elements (HRE) and associate to form a DNA loop. A transcription factor (TF) binds to the loop. *b*, Model for receptor oligomers binding to both HREs and forming a DNA loop. Contact between this receptor complex and TF enhances the binding of TF to the promoter region (P) or TF activity. (Arrow: direction of transcription.) Indirect evidence favours model *b* (see text).

molecules could be observed (Fig. 3c) and measurement indicated that these structures were formed by contacts between receptors rather than by a single receptor oligomer (not shown). Identical structures were seen with DNA fragments of the uteroglobin gene carrying only one HRE (Fig. 3d-f). The precise measurement of the apparent length of DNA before and after binding receptor was possible with these short DNA fragments; no significant differences were found, suggesting that no complex DNA structures were generated by binding to receptor.

At present there is no direct experimental evidence relating receptor-receptor interaction to hormone function. It should be noted, however, that in nearly all steroid-regulated genes there are several receptor binding sites^{5,7,8,10,11} which could facilitate the association of receptor oligomers. Moreover, the presence of both receptor binding sites upstream from the MMTV^{12,13} and chicken lysozyme¹⁴ genes is necessary for optimal hormonal activity. Loop formation may allow binding of transcription factor(s) on the loop DNA (Fig. 4a) or the loop may not itself be important but simply be a consequence of receptor-receptor interaction (see Fig. 1, c-f), which in turn modulates the activity of a transcription factor (Fig. 4b). This could explain the location of HREs at a distance from the promoter region of the gene. Cordingley *et al.*¹⁵ have recently shown that glucocorticoid receptor binding to the HREs in MMTV is accompanied by the binding of two proteins in the promoter region, one of which is the nuclear factor I. It is not known whether steroid receptor binding to DNA is a cooperative process when several HREs are present on the same DNA molecule.

Several steroid receptor genes have now been cloned¹⁶⁻¹⁸ which allows *in vitro* mutagenesis to locate the inter-receptor binding regions and correlate these with biological activity.

We thank Dr M. Crépin (Institut Pasteur, Paris) for the plasmid-containing MMTV DNA and J. F. Savouret for a mutant of the upstream region of the uteroglobin gene. We also thank N. Malpoint and M. Messina for typing the manuscript and Lorna Saint-Ange for assistance with revision. B. Théveny received a grant from Association pour la Recherche sur le Cancer.

Received 9 March; accepted 6 July 1987.

1. Parker, M. *Nature* **304**, 687-688 (1983).
2. Yamamoto, K. R. *Rev. Genet.* **19**, 209-252 (1985).
3. Ptashne, M. *Nature* **322**, 697-701 (1986).
4. Griffith, J., Hochschild, A. & Ptashne, M. *Nature* **322**, 750-752 (1986).
5. Bailly, A., Le Page, C., Rauch, M. & Milgrom, E. *EMBO J.* **5**, 3235-3241 (1986).
6. Logeat, F. *et al. Biochemistry* **24**, 1029-1035 (1985).
7. Scheidereit, C., Geisse, S., Westphal, H. M. & Beato, M. *Nature* **304**, 749-752 (1983).
8. Payvar, F. *et al. Cell* **35**, 381-392 (1983).
9. von der Ahe, D. *et al. Nature* **313**, 706-709 (1985).
10. Scheidereit, C. *et al. J. Steroid Biochem.* **24**, 19-24 (1986).
11. Becker, P. B., Gloss, B., Schmid, W., Strahle, U. & Schutz, G. *Nature* **324**, 686-688 (1986).
12. Kuhnle, B., Buetti, E. & Diggelmann, H. *J. molec. Biol.* **190**, 367-378 (1986).
13. Payvar, F. C. B., Miksic, R., Schutz, G., Arnemann, J. & Beato, M. *EMBO J.* **5**, 2237-2240 (1986).
14. Renkawitz, R., Schutz, G., von der Ahe, D. & Beato, M. *Cell* **37**, 503-510 (1984).
15. Cordingley, M. G., Tate Riegel, A. & Hager, G. L. *Cell* **48**, 261-270 (1987).
16. Weinberger, C. *et al. Science* **228**, 740-742 (1985).
17. Loosfelt, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 9045-9049 (1986).
18. Green, S. & Chambon, P. *Nature* **325**, 75-78 (1987).
19. Dubochet, J., Ducommun, M., Zollinger, M. & Kellenberger, E. *J. Ultrastruct. Res.* **35**, 147-163 (1971).
20. Revet, B., Zarling, D. A., Jovin, T. M. & Delain, E. *EMBO J.* **3**, 3353-3358 (1984).
21. Revet, B., Delain, E. & Bauer, R. *J. electron. Microsc.* **35**, Suppl., 2409-2410 (1986).

Heat shock factor is regulated differently in yeast and HeLa cells

Peter K. Sorger, Michael J. Lewis & Hugh R. B. Pelham

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

When cells are exposed to elevated temperatures, transcription of a small set of genes, the heat-shock genes, is activated¹⁻⁵. This response is mediated by a short DNA sequence, the heat-shock element (HSE)⁵⁻⁷, which is thought to be the binding site for a specific transcription factor⁸⁻¹². Studies with *Drosophila* show that this protein binds to HSEs only in heat-shocked cells, suggesting that changes in factor binding are responsible for gene activation¹⁰. We have investigated the properties of HSE-binding proteins from yeast and HeLa cells. In HeLa cells, binding activity is present only after heat shock. In contrast, control and heat-shocked yeast cells yield the same amount of HSE-binding activity; however, the mobility of protein-HSE complexes on polyacrylamide gels is altered following heat shock. This mobility difference can be significantly reduced by treatment of crude extracts with phosphatase. We propose that the yeast heat-shock factor binds constitutively to DNA but only activates transcription after heat-induced phosphorylation.

We looked for proteins in HeLa cells and yeast (*Saccharomyces cerevisiae*) with the ability to bind to a synthetic HSE sequence^{5,7}. This HSE matches a consensus derived from a variety of eukaryotic heat-shock promoters, and is functional in mammalian⁷, *Xenopus*⁷, *Drosophila*¹³ and yeast¹² cells. Figure 1 shows the results of binding assays using a probe that contains two overlapping consensus HSEs (this appears to be an optimal binding site *in vivo*)⁵. Binding proteins were detected by their ability to form complexes with labelled DNA and thus retard its migration on polyacrylamide gels^{14,15}. Heat-shocked HeLa cells contained such a protein, and binding could be abolished

by competition with excess unlabelled probe (Fig. 1a). With a small amount of extract two labelled bands were visible, but at higher protein concentrations only one slowly migrating band was seen (Fig. 1b).

Yeast extracts contained a similar activity, and again binding could be abolished by addition of excess unlabelled probe to the binding reaction (Fig. 1a). To compare the sequence specificities of the HeLa and yeast binding proteins, we performed binding assays in the presence of various unlabelled competitor DNAs: a monomeric HSE, a tetrameric HSE, and a dimeric HSE with a mutated sequence that matches the consensus in only six out of eight positions^{7,12}. Both yeast and HeLa proteins bound the tetramer more strongly than the monomer, and the mutant HSE (HSE12) was the weakest competitor of all (Fig. 1c). Competition by HSE12 was detected with the HeLa extract but not with the yeast extract. This correlates with the ability of HSE12 to support transcription in heat-shocked cells: it is active in mammalian cells⁷, but completely inactive in yeast¹². As a further comparison of the yeast and HeLa proteins, we performed DNase I protection experiments using a longer probe containing the synthetic dimeric HSE. As shown in Fig. 2, identical footprints were obtained with each extract. These experiments establish that yeast and HeLa cells contain HSE-binding proteins with similar sequence specificities.

We next examined the effect of heat shock on the binding activity. Whereas HeLa cells heat shocked to 45 °C for 30 min contained readily detectable levels of a specific HSE-binding protein (Fig. 1b), no activity could be detected in control cells. This inducibility supports the idea that the HSE-binding activity is involved in the regulation of heat shock genes. Induction probably involves modification of an existing protein, because it is not blocked by cycloheximide (data not shown). Similar results have recently been described by Kingston *et al.*¹⁶.

Different results were obtained when control and heat-shocked yeast cells were compared. Cells grown at 23 °C contained high levels of binding activity, and its amount did not change when the cells were heat shocked for 20 min at temperatures ranging from 34 °C to 39 °C (Fig. 3a). Moreover, the

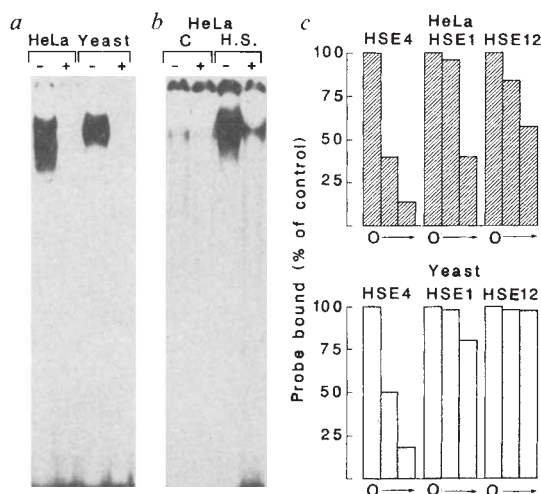


Fig. 1. HSE-binding proteins in HeLa cells and yeast. *a*, Gel electrophoresis of complexes formed with a dimeric HSE probe¹² and extracts from either heat-shocked HeLa cells or yeast. Assays were performed in the absence (–) or presence (+) of a 100-fold molar excess of unlabelled probe. Free probe has migrated off the bottom of the gel. *b*, Inducibility of binding activity in HeLa cells. Extracts²² (containing twice as much protein as in *a*) from control (C) and heat-shocked cells (H.S.; 45 min at 45 °C) were assayed with (+) or without (–) unlabelled competitor. *c*, Specificity of binding. Extracts from HeLa cells (top) and yeast (bottom) were incubated with labelled dimeric HSE probe (containing the sequence CTAGAAGCTTCTAGAAGCTTCTAG, the HSE consensus being CNNGAANNNTTCNNG), in the absence (0) or presence of increasing quantities (arrows) of unlabelled competitor DNA. The competitor contained either a single 8/8 match to the consensus HSE (HSE1, CTAGAAGCTTCTAG), four overlapping copies of this sequence (HSE4, CTAG(AAGCTTCTAG)₄) or two overlapping copies of a 6/8 match to the consensus (HSE12, CTAGAGATCTCTAGAGATCTCTAG)¹². The amount of bound probe was quantified by counting appropriate regions excised from gels. For HeLa extracts, successive bars correspond to a competitor: probe molar ratio of 0, 4 and 20; for yeast extracts the ratios were 0, 10 and 50.

Methods. Yeast extracts were prepared from the multiply protease-deficient strain BJ2168 (*leu2 trp1 ura3-52 prb1-1122 pep4-3 prc1-407 gal2*; a gift from D. Shore) and fractionated on heparin-agarose and calf thymus DNA-sepharose as described elsewhere¹². Binding reactions (20 µl) contained up to 40 µg of extract protein, 0.5 ng ³²P-end-labelled 62-bp probe, 1 µg supercoiled pUC19 DNA, 1 µg poly[d(I-C)], 20 mM HEPES pH 7.9, 1 mM MgCl₂, 60 mM KCl, 12% glycerol, 0.1% NP40 and 1 mM dithiothreitol. They were incubated for 30 min at room temperature and analysed on 4% polyacrylamide gels run in 22.5 mM Tris-borate, 0.6 mM EDTA. For competition experiments, the total amount of plasmid DNA was kept constant.

stability of DNA-protein complexes formed with extracts of control and heat-shocked cells, as estimated from their dissociation rates in the presence of excess unlabelled probe, was identical. With a monomeric HSE probe, the half-life of the complexes was ~20 min (Fig. 3c); complexes formed with a dimeric HSE in both control and heat-shocked extracts were more stable, and no dissociation was detected within two hours.

These results could be explained if our conditions for growing and harvesting the cells resulted in accidental heat-shock. Alternatively, the synthetic HSE might have some special property that causes it to be constitutively active. To test these possibilities, we introduced into yeast an expression plasmid in which the dimeric HSE sequence used in our binding assays lies upstream of a *CYC1-lacZ* fusion gene (this construct lacks the normal *CYC1* upstream activating sequences). The activity of the promoter was then monitored by β -galactosidase assay.

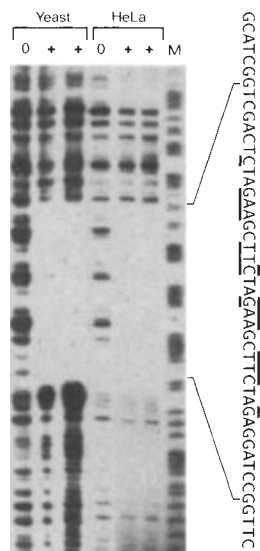


Fig. 2. Footprinting with yeast and HeLa extracts. The probe was derived from the plasmid pTKS2 (ref. 7). It was mixed with poly[d(I-C)] as carrier, and digested with DNase I after incubation without (0) or with (+) the yeast and HeLa extracts used in Fig. 1. M, marker track (the probe partially cleaved at purine residues). The DNA sequence is shown at the right, with the two overlapping HSE consensus sequences indicated by lines.

Very little activity was detected in cells grown at 23 °C but after heat shock for 20 min at temperatures between 34 °C and 39 °C there was a clear temperature-dependent induction of activity (Fig. 3a). When cells grown at 23 °C were centrifuged and resuspended as though for extract preparation, and then returned to growth medium, there was no induction of β -galactosidase activity (Fig. 3b). These results show that the synthetic HSE is functional in yeast and that despite its ability to form very stable protein-DNA complexes *in vitro*, it does not activate transcription constitutively. Furthermore, the presence of HSE-binding activity in control cells cannot be explained by accidental heat shock.

We conclude that the yeast heat-shock factor can bind to DNA whether or not cells are stressed. This implies that its ability to activate transcription is regulated independently of its ability to bind DNA. To see whether we could detect any heat-induced alterations in the factor, we examined the electrophoretic mobility of protein-HSE complexes formed in control and heat-shock extracts. Figure 4 shows examples of gels run for longer than normal, to accentuate possible differences. Two bands are visible in each lane, a major and a minor, more slowly migrating one. Extracts made from cells incubated at 39 °C for 20 min yielded complexes whose mobility was clearly reduced relative to those from control extracts (compare lanes 13 and 14). This effect could not be produced by heat shock of the control extract *in vitro* (lane 6), which rules out simple denaturation of the binding protein as its cause.

Yeast cells respond very rapidly to heat shock, but the response is transient, lasting no more than an hour¹⁷. Similarly, the mobility shift was detectable within three minutes of exposure to 39 °C (Fig. 4a) and was reversed within 30 min of recovery at 23 °C (Fig. 4c). The change in mobility was clearly progressive, during both induction of the response and recovery. The temperature required to induce a change in the protein-HSE complexes (Fig. 4b) also correlated well with the activation of heat shock transcription (Fig. 3a): the greatest change in mobility was observed at 39 °C, a partial effect at 36 °C, and a small effect at 33 °C. These results indicate that the changes detected by the gel assay are intimately connected with heat-shock gene activation.

Fig. 3. Detection of yeast HSE-binding activity does not require induction of the heat shock response. *a*, The β -galactosidase and HSE-binding activity in cells transformed with an HSE-*cyc1-lacZ* fusion gene. Cells were heat-shocked for 20 min at the indicated temperatures, and allowed to recover for 1 h at 23 °C. The β -galactosidase activity (●) was determined as described previously²³. Similarly treated cultures were harvested immediately after heat shock and assayed for HSE-binding activity. Assays were performed on duplicate cultures; the mean and range are shown (○). *b*, Transformed yeast cells were grown at 23 °C and β -galactosidase activities determined as in *a*. Values shown are for control cells (C), cells heat shocked at 39 °C (H.S.), and cells grown at 23 °C, harvested as though for determination of HSE-binding activity (see below), but resuspended in medium instead of being frozen in liquid N₂. After further incubation for 1 h at 23 °C, these mock-harvested cells (M) were assayed for β -galactosidase activity by the standard protocol. *c*, Stability of HSE-protein complexes. Binding reactions containing monomeric HSE probe and either heat shocked (●) or control (○) yeast extracts were incubated for 15 min before addition of a large excess of unlabelled competitor DNA (tetrameric HSE). After the indicated times, remaining complexes were assayed by gel electrophoresis.

Methods. The dimeric HSE sequence was inserted upstream of the TATA box of a disabled *cyc1-lacZ* fusion gene in the vector pLGA-178 (which contains the URA3 gene)²⁴, and the plasmid introduced into the strain BJ2168. Transformed cells were grown to early log phase in selective medium at 23 °C before assay. For determination of HSE-binding activity, cells were harvested by centrifugation, resuspended in 200 mM Tris-HCl pH 8.0, 10 mM MgCl₂, 10% glycerol, 1 mM PMSF, 0.5 mM TPCK, 0.025 mM TLCK, 2 μ g ml⁻¹ pepstatin A, rapidly frozen in liquid nitrogen, and ground in a mortar whilst still frozen. After thawing and centrifugation in a microfuge to remove debris, the extracts were assayed as in Fig. 1.

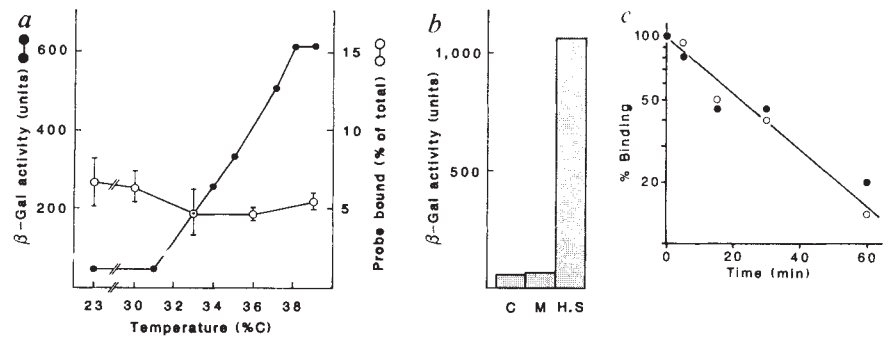
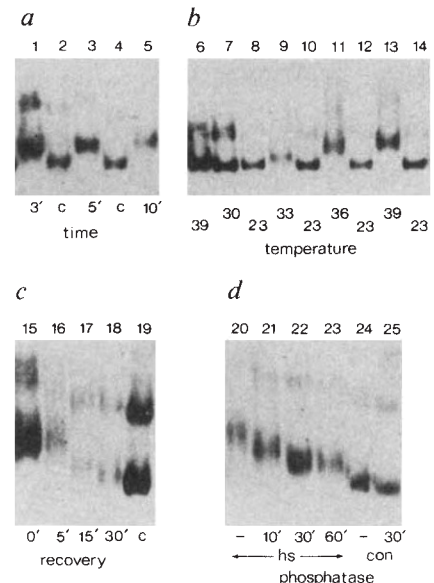


Fig. 4. Effect of heat shock on the mobility of protein-HSE complexes from yeast. Assays were performed as in Fig. 1, except that gels were run for 4 h (*a* and *b*) or 6 h (*c* and *d*) rather than 1 h. *a*, Time course of induction. Cells were grown at 23 °C, then incubated at 39 °C for the indicated number of minutes before harvesting. Lanes 2 and 4 are unshocked controls. *b*, Effect of heat shock temperature. Cells were heat shocked for 20 min at the indicated temperature (lanes 9, 11, 13) grown at 30 °C (lane 7) or left at 23 °C (lanes 8, 10, 12, 14). Lane 6 contains control extract that was heated to 39 °C for 20 min before assay. *c*, Recovery *in vivo*. Cells were heat shocked for 20 min at 39 °C and allowed to recover in an orbital shaker at 23 °C for the indicated times before harvesting; lane 19 shows an unshocked control. *d*, phosphatase treatment. Extracts containing 120 μ g of total protein from heat shocked cells (hs; 20 min at 39 °C) or control cells (con) were incubated at 37 °C for 30 min without phosphatase (lanes 20, 24) or for the indicated times with 20 units of calf intestinal phosphatase (Boehringer Mannheim) before assay.



What causes the change in mobility? A likely candidate is phosphorylation of the binding protein, and to test this we added calf intestinal phosphatase to the extracts. This had no effect on the complexes from control extracts (Fig. 4*d*, lanes 24 and 25), but progressively increased the mobility of the complexes formed in heat-shocked extracts (lanes 20-23). In the experiment shown in Fig. 4*d*, a residual mobility difference remained even after prolonged phosphatase treatment. This was frequently the case, although with at least one extract an apparently complete recovery was observed. A likely explanation of these results is that heat shock causes phosphorylation of the HSE-binding protein at multiple sites, and that some phosphates are removed from the protein very inefficiently. Nevertheless, we cannot rule out the possibility that an additional form of modification also occurs.

We have recently purified a 150K HSE-binding protein from unshocked yeast, and shown that it accounts for all the complexes seen when crude extracts are analysed by gel electrophoresis¹². Presumably, therefore, it is this protein that is phosphorylated upon heat shock; direct proof of this awaits isolation of the protein from heat-shocked cells.

The conclusion from this work is that the mechanism of

heat-shock gene activation differs between yeast and HeLa cells, even though the regulatory sequences are highly conserved. In HeLa cells, the affinity of the heat-shock factor for DNA is increased by heat shock, and binding presumably activates transcription. In yeast, the affinity of the factor for DNA does not change, and it must therefore be the nature of its interaction with other components of the transcriptional machinery that is altered. This does not exclude the possibility that the factors are modified by a common mechanism. Our results indicate that the yeast factor is modified by phosphorylation. So far, however, we have not been able to observe any effect of phosphatase on the HeLa cell factor.

In several instances, activation of transcription has been correlated with a change in the binding of putative transcription factors to DNA^{18,19}. In other cases, this correlation does not hold—for example, a protein binding the *c-fos* serum response element can be detected in both starved and serum-fed cells, although its activity is presumably modulated *in vivo*^{20,21}. The yeast heat-shock response may provide a precedent for the regulation of gene activity by phosphorylation of a constitutively bound factor.

We thank Gustav Ammerer, David Shore and Richard

Treisman for advice and materials, and our colleagues at the L.M.B. for many helpful discussions. P.K.S. was supported by a Marshall Commission Scholarship.

Note added in proof: We have recently purified the HSE-binding protein from heat-shocked yeast and shown that it is phosphorylated.

Received 1 May; accepted 21 July 1987.

1. Nover, L. (ed.) *Heat shock response of eukaryotic cells* (Springer, Berlin, 1984).
2. Craig, E. A. *CRC Crit. Rev. Biochem.* **18**, 239-280 (1985).
3. Lindquist, S. A. *Rev. Biochem.* **55**, 1151-1191 (1986).
4. Pelham, H. R. B. *Trends Genet.* **1**, 31-35 (1985).
5. Bienz, M. & Pelham, H. R. B. in *Advances in Genetics* (in the press).
6. Pelham, H. R. B. *Cell* **30**, 517-528 (1982).
7. Pelham, H. R. B. & Bienz, M. *EMBO J.* **1**, 1473-1477 (1982).
8. Parker, C. S. & Topol, J. *Cell* **37**, 273-283 (1984).
9. Wiederrecht, G., Shuey, D. J., Kibbe, W. A. & Parker, C. S. *Cell* **48**, 507-515 (1987).
10. Wu, C. *Nature* **309**, 229-234 (1984).
11. Wu, C. *Nature* **311**, 81-84 (1984).
12. Sorger, P. K. & Pelham, H. R. B. *EMBO J.* (in the press).
13. Berger, E. M., Torrey, D. & Morganelli, C. *Somat. Cell molec. Genet.* **12**, 433-440 (1986).
14. Fried, M. & Crothers, D. M. *Nucleic Acids Res.* **9**, 6505-6525 (1981).
15. Garner, M. M. & Revzin, A. *Nucleic Acids Res.* **9**, 3047-3060 (1981).
16. Kingston, R. E., Schuetz, T. J. & Larin, Z. *Molec. cell. Biol.* **7**, 1530-1534 (1987).
17. Slater, M. R. & Craig, E. A. *Molec. cell. Biol.* **7**, 1906-1916 (1987).
18. Sen, R. & Baltimore, D. *Cell* **47**, 921-928 (1986).
19. Seguin, C. & Hamer, D. H. *Science* **235**, 1383-1387 (1987).
20. Treisman, R. *Cell* **46**, 567-574 (1986).
21. Gilman, M. Z., Wilson, R. N. & Weinberg, R. A. *Molec. cell. Biol.* **6**, 4305-4316 (1986).
22. Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475-1489 (1983).
23. Breeden, L. & Nasmyth, K. *Cell* **48**, 389-397 (1987).
24. Guarente, L. & Mason, T. *Cell* **32**, 1279-1286 (1983).

A bacterial calcium-binding protein homologous to calmodulin

David G. Swan, Richard S. Hale, Namrita Dhillon & Peter F. Leadlay

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

Many of the effects of calcium ions in eukaryotic cells are mediated by calcium-binding regulatory proteins such as calmodulin, in which each calcium-binding site has a distinctive helix-loop-helix conformation termed the EF hand^{1,2}. Protein S from the spore coat of the Gram-negative bacterium *Myxococcus xanthus* has been shown to resemble calmodulin in its internally-duplicated structure and ability to bind calcium³. However, it has a β -sheet secondary structure^{4,5} rather than the helix-loop-helix arrangement of the eukaryotic proteins. We have determined the complete amino-acid sequence of a calcium-binding protein⁶ from the Gram-positive bacterium "*Streptomyces erythraeus*" by cloning and sequencing the corresponding gene. It contains four EF-hand motifs bearing remarkable sequence similarity to the calcium-binding sites in calmodulin. This implies that the EF-hand superfamily may have evolved from ancient proteins present in prokaryotes.

Calmodulin is the most prominent member of a family of small, acidic calcium-binding proteins, found in eukaryotic cells, through which calcium regulates a variety of cellular processes. The sequence similarity between these proteins is fairly extensive and is centred on the calcium-binding sites, of which there are four in calmodulin. Each calcium-binding site consists of a sequence of about 30 amino acids in which α -helical segments flank a 12-residue calcium-binding loop. Kretsinger has termed this structural motif the EF hand^{1,2}. Calmodulin has an established role in the activation of a number of enzymes involved in cellular regulation (see refs 7-9 for reviews). In contrast, the role of calcium-binding proteins in prokaryotes is obscure, although calmodulin-like activity has been reported in

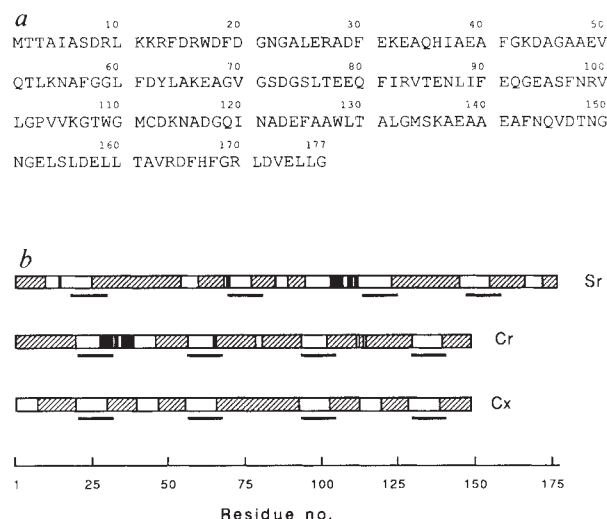


Fig. 1 "*S. erythraeus*"²⁶ calcium-binding protein: amino-acid sequence deduced from the sequence of the structural gene, and predicted secondary structure. **a**, The four potential calcium-binding loops span residues 20-31, 69-80, 113-124 and 147-158 respectively. **b**, The Robson¹⁴ algorithm was used to predict the secondary structure of "*S. erythraeus*" calcium-binding protein (Sr) and human calmodulin (Cr). The secondary structure of human calmodulin as deduced from X-ray crystallography¹⁵ is also shown (Cx). Hatched boxes, α -helical segments; filled boxes, β -sheet; unfilled boxes, β -turn or random coil conformations. Potential calcium-binding loops are underlined.

Methods. The structural gene for the calcium-binding protein from "*S. erythraeus*"²⁶ NRRL 2338 was detected, in chromosomal DNA cut with *Kpn*I, using oligonucleotide probes designed from a knowledge of part of the amino-acid sequence, and cloned into the multicopy vector pUC18 in *E. coli*. A 0.8-kilobase (kb) DNA fragment containing the entire structural gene was sequenced using the Sanger dideoxy chain termination method²⁷. Details of the cloning, the complete DNA sequence of this fragment and criteria for identifying the reading frame will be described elsewhere.

*Escherichia coli*¹⁰ and *Bacillus subtilis*¹¹. When a small, acidic protein from the erythromycin-producing bacterium "*Streptomyces erythraeus*" was found to contain an EF hand motif and was shown to bind calcium⁶, it led us to examine the primary structure of this protein in more detail.

We have isolated and sequenced the structural gene for the calcium-binding protein from "*S. erythraeus*" and deduced the complete amino-acid sequence of the protein (Fig. 1a). The protein contains 177 amino acids (relative molecular mass (M_r) 20,090), compared to the 148 amino acids in human calmodulin¹² (M_r 16,723). Like calmodulin, the "*S. erythraeus*" protein contains a high percentage of negatively charged amino acids, but the overall amino-acid composition is distinctly different.

The amino-acid sequences of the "*S. erythraeus*" calcium-binding protein and of human calmodulin were compared using the Diagon algorithm of Staden¹³, in which two protein sequences are displayed along the x and y axes, respectively, and scored for similarity. Several regions of similarity were revealed (Fig. 2), and these corresponded exactly to the positions of the potential calcium-binding sites in both proteins. The pattern is not perfectly regular, because the calcium-binding sites in the bacterial protein are not as evenly-spaced along the polypeptide chain as they are in calmodulin. Calcium-binding site II in the "*S. erythraeus*" protein (residues 113-124) is only revealed, at this level of stringency, by alignment with one out of the four calcium-binding sites in calmodulin. The amino-acid sequence in the N-terminal part of the presumed calcium-binding loop

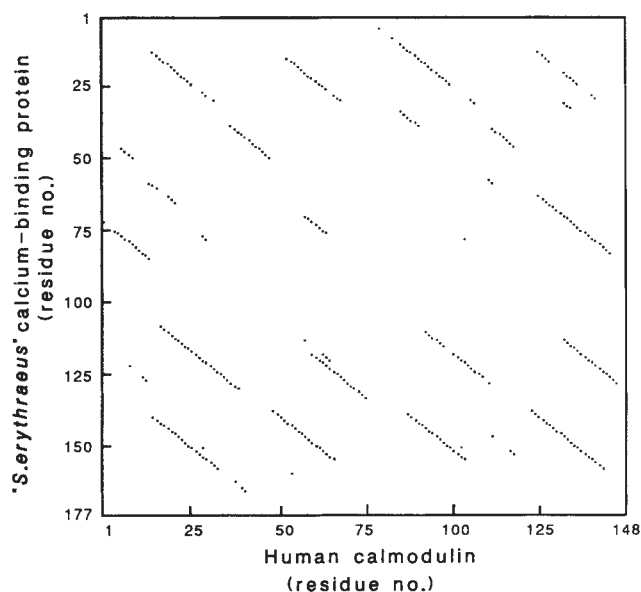


Fig. 2 Comparison of the amino-acid sequence of "*S. erythraeus*" calcium-binding protein with that of human calmodulin, using a Diagon plot¹³. The comparison used a 21-residue segment and generated a dot if the score exceeded 230 (ref. 13).

<i>a</i>	1	3	5	7	9	12						
I	D	F	D	G	N	G	A	L	E	R	A	D
II	G	V	G	S	D	G	S	L	T	E	E	Q
III	D	K	N	A	D	G	Q	I	N	A	D	E
IV	D	T	N	G	N	G	E	L	S	L	D	E

<i>b</i>	1	3	5	7	9	12						
I	D	K	D	G	D	G	T	I	T	T	K	E
II	D	A	D	G	N	G	T	I	D	F	P	E
III	D	K	D	G	N	G	Y	I	S	A	A	E
IV	D	I	D	G	D	G	Q	V	N	Y	E	E

Fig. 3 Alignment of the four potential calcium-binding loops, numbered from the N-terminus (I–IV). *a*, "*S. erythraeus*" calcium-binding protein; *b*, human calmodulin. The six oxygen ligands to Ca^{2+} are provided by residues 1, 3, 5, 7, 9 and 12 in each loop.

at this site is atypical (Fig. 3). In contrast to calmodulin, there is no overall sequence homology between the two halves of the protein.

The helix-loop-helix arrangement at each of the potential EF-hand structures in the "*S. erythraeus*" protein is predicted by the Robson algorithm¹⁴ (Fig. 1*b*). Although the actual secondary structure of calmodulin, as deduced from the X-ray crystal structure¹⁵, differs in important details from the structure predicted by the Robson algorithm (Fig. 1*b*), this analysis confirms that the calcium-binding protein from "*S. erythraeus*" has a structural organization similar to that found in calmodulins, rather than to the β -sheet structure found^{4,5} for *M. xanthus* protein S. The flanking α -helices at each EF hand site are proposed to be amphipathic². In agreement with this, although there is no evident sequence similarity, the residues predicted to lie on the hydrophobic side of the helices are almost uniformly non-polar in the "*S. erythraeus*" calcium-binding protein, just as in calmodulin.

The aligned sequences of the 12 residues in each of the four potential calcium-binding loops in the "*S. erythraeus*" protein

(Fig. 3) show the pattern of conserved residues anticipated² for an authentic EF-hand structure. Side-chains at positions 1, 3, 5, 9 and 12 (Fig. 3) provide oxygen-containing ligands to calcium, and these are predominantly acidic. At position 7, the ligand is provided by the carbonyl oxygen in the main polypeptide chain, and the side-chain at this position is not conserved either in calmodulin or in the bacterial protein. The glycine at position 6 and the aliphatic hydrophobic residue at position 8 are critical for calcium binding and are conserved. Site II of the "*S. erythraeus*" calcium-binding protein is unusual in having glycine at both positions 1 and 3. It may be that calcium does not bind to this site, but it is also possible that solvent water may provide the ligands at these two positions².

The EF-hand superfamily of calcium-binding proteins shows an astonishing diversity of function, based upon the same common structural motif. Kretsinger has made the generalization² that all EF-hand proteins are involved in calcium-dependent regulation. Some hybrid proteins^{16–18} contain EF hands and a recognizable catalytic site in the same polypeptide chain, but in general the cellular targets of calcium-modulated proteins must be inferred from *in vitro* studies. Recently, genetic experiments have suggested a function for calmodulin¹⁹ and for the *CDC31* gene product²⁰ in the control of the yeast cell cycle. Streptomycetes accumulate large amounts of calcium during differentiation and sporulation²¹ and calcium ions have also been implicated in initiation of the germination of streptomycete spores²². Based on the apparent structural homology to eukaryotic EF-hand calcium-binding proteins, it is tempting to ascribe a regulatory function to the "*S. erythraeus*" calcium-binding protein. Experiments are now in progress, using the cloned gene as well as the purified protein, to try to elucidate the biological function of this protein.

Our results suggest that the EF-hand motif may have arisen in an ancient protein, before the divergence of the eukaryotes and prokaryotes. The similarity of the bacterial protein to calmodulin, rather than to other known EF-hand proteins, supports the view²³ that the eukaryotic EF-hand superfamily has diverged from a common calmodulin-like ancestor with four calcium-binding domains. Less likely alternatives are that convergent evolution, or later gene transfer from eukaryote to prokaryote²⁴ may be involved. These findings should encourage the search for related proteins in other prokaryotes^{10,11,25}.

This work was supported by grants from the M.R.C. (to P.F.L. for R.S.H.) and the Wellcome Trust (to P.F.L. for D.S.). N.D. thanks the Nehru Commonwealth Trust for a Scholarship.

Received 28 May; accepted 16 July 1987.

- Kretsinger, R. H. & Nockolds, C. E. *J. biol. Chem.* **248**, 3313–3326 (1973).
- Kretsinger, R. H. *CRC crit. Rev. Biochem.* **8**, 119–174 (1980).
- Inouye, S., Franceschini, T. & Inouye, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6829–6833 (1983).
- Inouye, S., Harada, W., Zusman, D. & Inouye, M. *J. Bact.* **148**, 678–683 (1981).
- Wistow, G., Summers, L. & Blundell, T. L. *Nature* **315**, 771–773 (1985).
- Leadlay, P. F., Roberts, G. & Walker, J. E. *FEBS Lett.* **178**, 157–160 (1984).
- Cheung, W. Y. in *Calcium and Cell Function* Vol. 1 (ed. Cheung, W. Y.) 1–12 (Academic, New York, 1980).
- Klee, C. B. & Vanaman, T. C. *Adv. Protein Chem.* **35**, 213–321 (1982).
- Van Eldik, L. J. & Watterson, D. M. in *Calcium and Cell Physiology* (ed. Marme, D.) 105–126 (Springer, Berlin, 1985).
- Iwasa, Y., Yonemitsu, K., Matsui, K., Kukunaga, K. & Miyamoto, E. *Biochem. biophys. Res. Commun.* **98**, 656–660 (1981).
- Fry, L., Villa, I. J., Kuehn, G. D. & Hageman, J. H. *Biochem. biophys. Res. Commun.* **134**, 212–217 (1986).
- Sasagawa, T. *et al. Biochemistry* **21**, 2565–2569 (1982).
- Staden, R. *Nucleic Acids Res.* **10**, 2951–2961 (1982).
- Garnier, J., Osguthorpe, D. J. & Robson, B. *J. molec. Biol.* **120**, 97–120 (1978).
- Babu, Y. S. *et al. Nature* **315**, 37–41 (1985).
- Charbonneau, H. *et al. Biochemistry* **24**, 6762–6771 (1985).
- Ohno, S. *et al. Nature* **312**, 566–570 (1984).
- Parker, P. J. *et al. Science* **233**, 853–859 (1986).
- Davis, T. N., Urdea, M. S., Masiaz, F. R. & Thorner, J. *Cell* **47**, 423–431 (1986).
- Baum, P., Furlong, C. & Byers, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5512–5516 (1986).
- Salas, J. A., Guijarro, J. A. & Hardisson, C. *J. Bact.* **155**, 1316–1323 (1983).
- Eaton, D. & Ensign, J. C. *J. Bact.* **143**, 377–382 (1980).
- Goodman, M., Pechere, J.-F., Haiech, J. & Demaille, J. G. *J. molec. Evol.* **13**, 331–352 (1979).
- Carlson, T. A. & Chelon, B. K. *Nature* **322**, 568–570 (1986).
- Yates, L. & Vanaman, T. C. *Fedn Proc.* **40**, 1738 (1981).
- Labeda, D. P. *Int. J. system. Bact.* **37**, 19–22 (1987).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).

Site-directed mutation affecting polyomavirus capsid self-assembly *in vitro*

R. L. Garcea*, D. M. Salunke† & D. L. D. Caspar†

* Division of Pediatric Oncology, Dana-Farber Cancer Institute, The Children's Hospital, and the Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA
† Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254, USA

Nonequivalent bonding of identical protein subunits occurs in the polyomavirus capsid where identical pentameric capsomeres occupy both hexavalent and pentavalent positions in the icosahedral surface lattice¹. The polyomavirus major capsid protein VP1, purified after expression of the recombinant gene in *Escherichia coli*², has been isolated as capsomeres that self-assemble into capsid-like structures *in vitro*³. The ability to switch bonding specificity in different symmetry environments therefore must be intrinsic to the VP1 molecule. *In vitro* self-assembly provides an assay for VP1 mutations affecting capsomere and capsid formation. We report here that a directed mutation in the VP1 expression vector, leading to a protein truncated at the carboxy terminus, results in a mutant VP1 that forms capsomeres, but not capsids, in the *in vitro* assembly assay. The carboxy terminus of VP1 therefore appears to be involved in the specific bonding responsible for the non-equivalent association of capsomeres.

Efficient expression of polyomavirus VP1 in *E. coli* allows the characterization of this protein and the manipulation of the gene to introduce mutations for the study of capsid assembly. The VP1 expression vector pALVP1tac produces unmodified VP1 protein in *E. coli* RB791 under the inducible control of the hybrid *tac* promoter². Two unique *Nco*I restriction endonuclease sites near the carboxy terminus of VP1 were used to generate a deletion in the expressed VP1 protein (Fig. 1). The

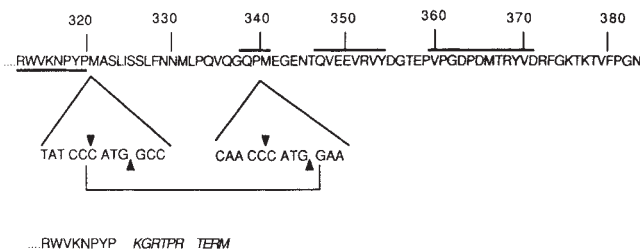


Fig. 1 Protein and nucleotide sequences at the carboxy terminus of VP1. The carboxy-terminal amino-acid sequence is designated by the single letter code. The sequence of VP1 from strain RA (kindly supplied by T. Benjamin), from which pALVP1tac was derived, is identical to strain A2 (ref. 6) in this region. Shown below the amino-acid sequence are the nucleotide sequences including the two *Nco*I restriction sites unique to the VP1 bacterial expression vector, pALVP1tac. A sample of 5 µg pALVP1tac was digested with *Nco*I, and the large DNA fragment was isolated from low-melting-point agarose (Seakem Seaplaque). This fragment was then digested with mung bean nuclease, and re-isolated. Approximately 1 µg of the now blunt-ended vector was self-ligated, and then transformed into *E. coli* RB791 (*lacI*^q). Transformants were screened by digestion with *Nco*I, to assay for loss of this restriction site, and by digestion with *Hinc*II, to assay for a 64 base-pair decrease in size of the appropriate restriction fragment. The modified vector shifts the reading frame of VP1, as indicated, by the ligation of CC to GAA. The result is a deletion of 63 amino acids from the wild-type sequence and substitution of 6 out-of-frame amino-acid residues (indicated by italics) before a termination codon (term) is reached. Amino acids identical to SV40 VP1 are indicated by a superscript bar.

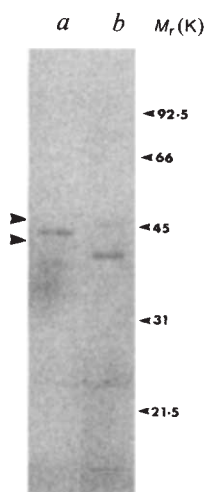


Fig. 2 SDS-PAGE of the purified mutant VP1 protein. *E. coli* RB791 containing the deleted expression vector were induced with IPTG and the mutant VP1 purified according to the protocol previously described for the wild-type protein². The purification properties of the mutant and wild-type proteins were similar. Shown are wild-type VP1 (a), and the purified mutant VP1 (b) after SDS-PAGE and Coomassie blue staining. VP1 migrates on the gel anomalously slower than expected for its known molecular weight. The shift in electrophoretic mobility of the mutant protein is consistent with the deletion of 57 amino acids. *M_r* markers are indicated.

mutant protein is 57 amino-acid residues shorter than the wild type.

Bacterial transformants carrying the deleted expression vector were induced with IPTG and the proteins were labelled with ³⁵S-methionine². Bacterial cell lysates were immunoprecipitated with anti-VP1 antibodies, and the presence of an immunoreactive protein of the expected relative molecular mass (*M_r*) 41,000 was identified by SDS-PAGE and autoradiography of the dried gel (data not shown). Purification of the mutant protein was carried out according to the protocol previously described for the wild-type VP1 protein². As shown in Fig. 2, this purification scheme was successful for the mutant protein, and the final yield was ~1 mg of apparently homogenous protein per liter of induced bacteria.

Velocity sedimentation analysis in a sucrose gradient of the mutant VP1 protein showed a single major component with a sedimentation coefficient of 7.5 S (Fig. 3), which is comparable to that of the wild-type pentameric capsomeres³. This value of the sedimentation coefficient is lower than expected for a compact capsomere, suggesting that local unfolding leads to an increase in frictional drag for both mutant and wild-type proteins.

Electron micrographs of the negatively-stained mutant protein show that it behaves like the wild type under conditions favouring dissociation of viral capsids, yielding a uniform population of capsomere-like aggregates (Fig. 4a, b). Many of these capsomeres are aligned on the grid with their stain-filled axis approximately normal to the surface and have a pentagonal shape. When transferred to conditions favouring association, namely removal of the disulphide reducing agent and readdition of calcium, the wild-type capsomeres self-assemble into capsid-like structures similar to native capsids in size, shape, and subunit packing (Fig. 4c). The mutant VP1 capsomeres, however, do not associate into capsid-like structures or other polymorphic aggregates under these conditions (Fig. 4d). Neither have we observed aggregation of the mutant capsomeres under conditions, such as high concentrations of ammonium sulphate, normally leading to polymorphic assembly of the wild-type capsomeres³. A functional property of the mutant protein which may be compared to wild type is haemagglutination. Wild-type recombinant VP1 capsomeres and capsid preparations haemagglutinate sheep erythrocytes at a protein to haemagglutinating antibody (HA) titre ratio equivalent to that of isolated virions⁴. Assay of the haemagglutination of sheep erythrocytes⁵ by the mutant capsomeres yielded a ratio of 10 ng protein per 1 ± 0.5 HA unit which is comparable to that for wild-type recombinant protein and purified virus.

The C-terminal deletion of VP1 described here does not significantly alter either the structure of the capsomere, as determined by morphology in the electron microscope and by sedimentation, or the specific functional property of haemag-

Fig. 3 SDS-PAGE analysis of fractions after sedimentation of the mutant VP1 protein in a sucrose gradient. The purified mutant VP1 protein (Fig. 2) was dialysed into 50 mM Tris pH 7.4, 15 mM 2-mercaptoethanol, 1 mM EDTA, and 250 mM NaCl, and layered on a 5–20% sucrose gradient formed in the same buffer. Samples were sedimented in an SW40 rotor (Beckman) at 39,000 r.p.m. for 16 h at 4 °C. Eighteen 500 μ l fractions were collected from below, 10 μ g bovine serum albumin added as carrier, trichloroacetic acid added to 10%, and the protein precipitated overnight at 4 °C. The precipitates were washed with 95% ethanol, dried, and subjected to SDS-PAGE in 10% gels. Sedimentation markers in a parallel gradient were β -galactosidase (19S), catalase (11.3S), and alkaline phosphatase (6.2S), identified by their enzymatic activity. The sedimentation coefficient of the mutant protein is approximately 7.5 S.

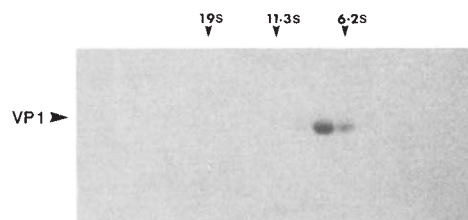
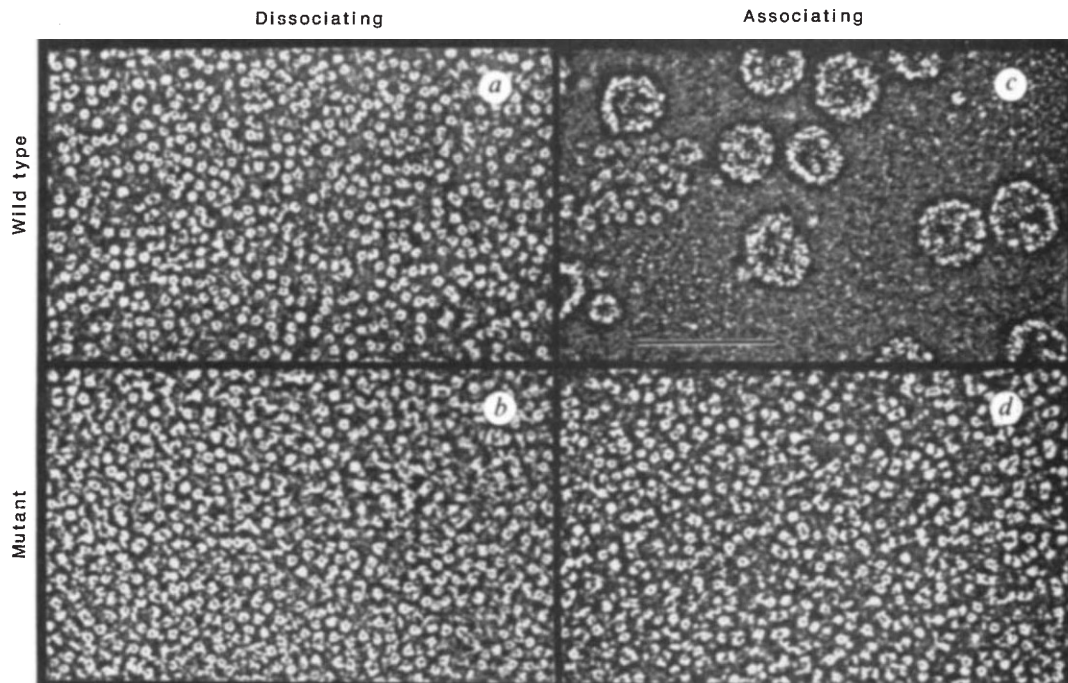


Fig. 4 Electron micrographs of wild-type and mutant VP1 proteins under dissociating and associating conditions.

a, Wild-type VP1 in 10 mM Tris pH 7.2, 15 mM 2-mercaptoethanol, 1 mM EDTA, 5% glycerol and 100 mM NaCl, is seen as separated capsomeres with their stain-filled axes oriented generally normal to the grid.

Native polyomavirus capsids dissociate into capsomeres under these solvent conditions¹¹. **b**, Mutant VP1 under the same dissociating conditions is seen as capsomeres which are indistinguishable from those of the wild-type protein.

c, Wild-type VP1 capsomeres in 10 mM Tris pH 7.2, 5% glycerol 0.5 mM CaCl₂, self-assemble into capsid-like structures which are similar to polyomavirus capsids. A puddle of capsomeres at the left appears to be derived from a capsid aggregate broken down on the grid. **d**, Mutant VP1 capsomeres under these associating conditions do not form capsids or other polymorphic aggregates. The mutant protein remains stable as unassociated capsomeres which are indistinguishable from those in either **a** or **b**. The micrographs have been computer processed to filter noise and enhance contrast. Scale bar, 1,000 Å.



glutination. The retention of these features implies that the protein segments responsible for them are not localized at the C terminus. Moreover, the inability of the mutant capsomeres to assemble into capsid-like structures is therefore unlikely to be attributable to a gross alteration in the three-dimensional structure of the protein.

The polyomavirus capsid is composed of 72 pentameric capsomeres of VP1 (ref. 1). From the electron density map of the virus capsid, the VP1 molecule appears to consist of two regions: a projecting domain forming the pentagonally symmetric protrusions, and a shell domain connecting the neighbouring capsomeres. Because capsomeres occupy both hexavalent and pentavalent positions on the $T = 7d$ icosahedral surface, the bonding interactions between capsomeres must be non-equivalent in the different symmetry positions. Thus, the bonding between capsomeres requires some coordinated mechanism for switching specificity which depends on capsomere position. The ability of recombinant VP1 protein to self-assemble into capsid-like structures implies that the switching is intrinsic to VP1 (ref. 3). Because the mutant protein can form capsomeres with a structure similar to the wild type but is unable to self-assemble into capsid structures, the C terminus of VP1 is implicated in the specific bonding between capsomeres. It is interesting to note the considerable sequence homology with SV40 VP1 in this region^{6,7} (Fig. 1), suggesting a conserved function. The C terminus may be involved in capsomere association by forming part of the bonding region itself, or by stabiliz-

ing the different bonding conformations between capsomeres. In either case, it is likely that the C terminus is localized in the shell domain where contacts between capsomeres occur.

Similar protein engineering methods^{8,9} have been used Katsura¹⁰ to study phage tail length determination *in vivo*. Supramolecular protein engineering¹⁰ can thus be applied in various ways to analyse higher levels of organization in structural biology. Further directed mutagenesis of the recombinant polyomavirus VP1 protein expression vector should delineate the protein segments involved in the bond switching between capsomeres and lead to the identification of other structural features significant in capsid assembly.

We thank Lynne Montross for technical assistance. This work was carried out with the support of grants from the American Cancer Society to R.L.G. and the National Cancer Institute to R.L.G. and D.L.D.C.

Received 21 May; accepted 6 July 1987.

1. Rayment, I., Baker, T. S., Caspar, D. L. D. & Murakami, W. T. *Nature* **295**, 110–115 (1982).
2. Leavitt, A. D., Roberts, T. M. & Garcea, R. L. *J. biol. Chem.* **260**, 12803–12809 (1985).
3. Salunke, D. M., Caspar, D. L. D. & Garcea, R. L. *Cell* **46**, 895–904 (1986).
4. Zullo, J., Stiles, C. D. & Garcea, R. L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1210–1214 (1986).
5. Garcea, R. L. & Benjamin, T. L. *Virology* **130**, 65–75 (1983).
6. Soeda, E., Arrand, J. R., Smolar, N., Walsh, J. E. & Griffin, B. E. *Nature* **283**, 445–453 (1980).
7. Toozé, J. in *DNA Tumor Viruses*, 2 edn, Appendix B 870–872 (Cold Spring Harbor, New York).
8. Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M. & Smith, M. *Nature* **299**, 756–758 (1982).
9. Ulmer, K. M. *Science* **219**, 666–671 (1983).
10. Katsura, I. *Nature* **327**, 73–75 (1987).
11. Brady, J. N., Winston, V. D. & Consigli, R. A. *J. Virology* **23**, 717–724 (1977).

GUIDE TO AUTHORS

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully.

Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

CATEGORIES OF PAPER

Review Articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50–80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

News and Views articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

Scientific Correspondence is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

GENERAL

All contributions submitted for publication in *Nature* should conform with these rules.

Manuscripts should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

Titles should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

Figure legends must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

References should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unrefereed abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963–65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

Footnotes should not be used except for changed addresses or to identify the corresponding author if different from the first-named.

Acknowledgements should be brief. Grant numbers and contribution numbers are not allowed.

Submission. All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

Novelties from software to services

This week's New on the Market is a combination of laboratory conveniences and advances, including a snap-in database and systems for keeping track of chemicals, animals and chromosomes.

DNASTar has taken another step forward in producing **software for molecular biologists**. Now its comprehensive molecular biology software, plus the complete GenBank and NBFR-PIR databases, are available on a compact disk, called LaserGene (*Reader Service No. 100*). LaserGene provides random access to 552 Mb of data — the equivalent of 1,444 floppy diskettes. DNASTar says the compact disk technology uses sophisticated error correction protocols resulting in less than one error per 2,000 disks. When used in conjunction with the DNASTar system, full database searches take only seconds. Updates of GenBank and NBFR-PIR databases are released quarterly, and are installed by simply replacing the disk.

Rotofor is the name of the newest **preparative scale electrophoresis cell** from Bio-Rad (*Reader Service No. 101*). The Rotofor is a free-solution isoelectric focusing apparatus. It uses a cylindrical



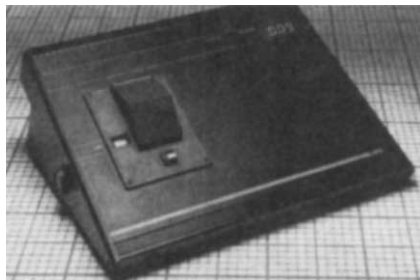
Eliminates mixing problems with preparative IEF techniques.

focusing chamber that rotates around its axis to overcome the problems inherent in scaling up IEF techniques. Nineteen parallel, monofilament polyester membrane screens divide the cylindrical chamber into 20 discrete compartments, allowing the solutions to be rapidly and easily collected after the proteins are focused. Bio-Rad says the Rotofor is capable of achieving greater than 10-fold purification of a protein sample in a single 4-hour run, with recovery of proteins of single-band purity in certain cases.

Charles River Biotechnical Services is now performing **safety testing of cell lines** and products for use in human therapy, in accordance with US Food and Drug Administration and European Economic Community regulations (*Reader Service No. 102*). The tests include detection of 16 murine viruses, XC plaque and S⁺ L⁻ murine retrovirus assays, electron microscopy services, DNA hybridization and DNA transformation assays, and tests for EBV and human viruses in cell culture. CRBS also offers production services for monoclonal antibodies or other biomol-

ecules using *in vivo* or Opticell *in vitro* techniques, under GMP or GLP conditions.

If the wedge is the shape of things to come, then Walden Precision Apparatus is on the cutting edge with its **S110 spectrophotometer** (*Reader Service No. 103*). The



The S110: a new angle on spectrophotometry.

instrument is designed to eliminate user adjustment and zeroing. An autozeroing device readies the instrument after the reference cuvette is placed in one slot of the two-position cuvette holder. The only controls provided are the power switch, wavelength selector and cuvette change-over. The £825 (UK) spectrophotometer gives sample absorbance from 0.000 to 1.999 A, using wavelengths between 320 and 900 nm.

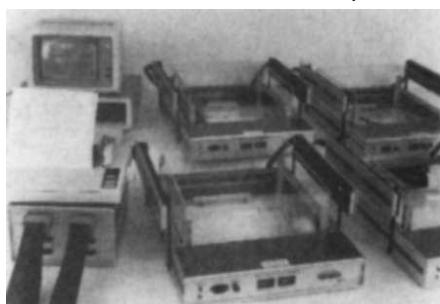
Custom **transgenic mice** are available from Clontech Laboratories, for use in medical research and pharmaceutical development (*Reader Service No. 104*). Clontech will tailor the gene supplied, fuse it to an appropriate promoter and microinject it into the germline of the chosen mouse strain. Clontech guarantees that first generation transgenic mice will be positive for the integration of the DNA, as shown by Southern blot analysis, and will protect all gene constructs and transgenic animals as proprietary products for use solely by the customer.

Often it is vitally important to keep track of hazardous substances, and document their uses and disposal. Fisher has a **hazardous substance tracking system** called Trace to help laboratories account for hazardous materials, even if they are subdivided into several containers (*Reader Service No. 105*). Trace is a software program based on the Unisys 500 minicomputer that assigns a unique identifier to each container instead of each chemical, and prints bar-coded labels for them. Each container record includes the destination area of the substance, the owner ID, the expiration date, container type, lot number, date and time received, and comments of up to 25 characters. The

system can be purchased directly for \$15,000–60,000 (US), or as a part of a contract arrangement for Fisher equipment or supplies.

An **automated chromosome analysis system**, Cytoscan, is available from Image Recognition Systems (*Reader Service No. 106*). The system is designed for high-speed metaphase finding and karyotyping in cytogenetics, to assist in the diagnosis of chromosome defects, but can also aid detection of chromosomal abnormalities due to radiation or other agents. The \$180,000–225,000 (US) Cytoscan consists of a computer-controlled microscope and a colour monitor. Images may be manipulated via the keyboard or an optical mouse that can act as "electronic scissors" to split overlapping and touching chromosomes. Karyograms can be printed out, or stored permanently on the system.

The **animal activity meter** system from Columbus Instruments International can be used to record the positions of animals in 16 cages 10 times per second, according to the company (*Reader Service No. 107*). The Columbus Auto-Track animal activity meter has a new interface for IBM PC computers that allows data from up to 16 animal activity sensors to be collected and evaluated. The animal activity sensors



Tracks animals in up to 16 cages simultaneously.

operate on an infrared beam principle, with 15×15 infrared beams in X and Y coordinates, and cover a cage size of 37×37 cm. Two vertical sensors can be positioned on two adjustable levels to allow the distinction of animal rearings and jumps. Computer files created with the \$16,290 (US) Auto-Track system, including an animal activity sensor, interface and computer, can be used with Lotus 1-2-3, Stat-Pro, and other statistical programs.

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

Student aspirations decidedly financial

Richard Pearson

Careers in high-technology industry and scientific research are not top of all graduates' lists, and college-leavers are taking a hard-headed approach to employment.

In theory, the ever-increasing investment in information technology in schools and the education system to make students computer literate should make life easier for high-tech recruiters. But in reality the reverse is true, as more and more college students in the United States are turning away from computer studies and towards careers in the professions. This shift in career preferences, when allied to the demographic downturn¹, makes gloomy reading for the high-tech industries and the research community. The one positive outcome of the latest freshman survey is the recent rise in student interest in careers in education.

The survey^{2,3}, which covers over 250,000 students in 550 colleges and universities across the United States, has now been running for 20 years and charts the movement from the heady optimistic days of the 1960s to the harsher more competitive climate of the 1980s. Over this period, the student population has grown by 40 per cent and the representation by women and minorities has grown even faster; in 1986 the proportion of women had risen to nearly 52 per cent. Black representation had nearly doubled to 9 per cent, although most of this increase had come in the 1960s and early 1970s. The evidence on entry standards is mixed, with students rating themselves higher in terms of personality and skill traits, and their high-school grades being considerably higher. The latter is, however, put down to grade inflation. College test scores have declined as have faculties' judgement of their students' abilities, especially their verbal skills. The one growth skill has been that of computer programming, 50 per cent now having this skill, up from 17 per cent in 1982 and 5 per cent in 1977.

The survey shows rising student dependence on student loans and part-time and summer working as federal aid and grants have been cut back. One consequence has been the squeezing out of students from low-income families; in 1980, 40 per cent of freshmen reported family incomes of under \$20,000 a year compared with only 20 per cent in 1986, with comparable falls for students with families earning under \$30,000. These are far higher rates of decline than can be accounted for by inflation. Heavy reliance on loans may contribute to the decision of some students to drop out of college, further emphasizing the growing divisions between the children from poorer and richer families.

What then are the students seeking? In the 1960s, "developing a meaningful philosophy of life" was a key objective for students; this is now of very low interest, while "being well off financially" is now endorsed as a priority by 71 per cent of students, up from 44 per cent 20 years ago. Other related values that have increased dramatically in importance relate to power and status — a key motivating factor in going to college in the 1980s is "to be able to make more money". Altruistic and social concerns have plunged, "helping others", "promoting racial understanding" and "cleaning up the environment" are now out of favour.

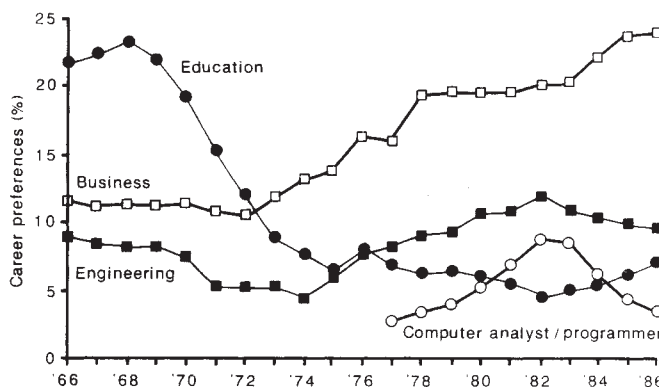
The main growth subject areas over the past 20 years have been business studies, which accounted for one in four students in 1986, up from less than one in six 20 years earlier, and engineering and computing, although interest has started to decline in the past few years (see figure). Originally these were male-dominated subjects. Women now have parity in business studies, but are still outnumbered 6 to 1 in engineering and 2 to 1 in computing. Interest in these latter two subjects peaked in 1983 and they seem

particularly sensitive to students' perceptions of job opportunities in these fields which suffered reverses over this period. However, the decline in student interest in computing seems to have preceded the shakeout in the computer industry. In the sciences, mathematics and statistics show a major decline from nearly 5 to under 1 per cent of freshmen over the 20 years. Other subjects showing major declines include both the physical and biological sciences, psychology and until recently education². The latter has been dominated by women but the decline has been equally severe for both sexes.

Subject and career choice are closely allied and there has been a doubling of the proportion of students intending to take up careers in business over the past 20 years, to account for 1 in 4 in 1986. Computing also grew rapidly to a peak of 9 per cent in 1982, but interest has since halved. Engineering likewise grew until 1982 and

has since fallen back. Education has also shown a massive decline but with the present shortages of teachers and improved pay and conditions, students are beginning to rediscover teaching as a career. The level of interest is still, however, far too low to accommodate the rapid increase in demand for new teachers that will be needed to overcome the retirement bulge in the 1990s. Another problem is the declining interest in doctoral studies, further compounding worries about shortages of faculty in the next decade. Other areas of declining interest include medicine, dentistry, nursing and social work, and interestingly, law.

The freshman surveys provide a valuable guide to changing student. It is clear that, for some subjects, labour market conditions have a direct and rapid effect on



Freshman career preferences for the years 1966–86. From ref. 3.

student aspirations, as is shown by the significant and rapid downturn in the case of computing. Unfortunately, this sensitivity relates to current conditions and not those likely to be prevailing 5 years from now when students will be job-hunting nor to longer-term career prospects. What is less clear is why only some subjects are susceptible to such influences. Is it that in certain potentially vocational subjects such as biology, no vocational linkages are made, or do the apparently 'vocational' subjects attract less committed students whose interests are in the more tangible rewards accruing to their studies?

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.

1. Pearson, R. *Nature* 326, 528 (1987).
2. *The American Freshman: Twenty Year Trends* (University of California, 1987).
3. *The American Freshman: National Norms for Fall 1986* (University of California, 1987).